PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 13 August 1999 (13.08.99)	
International application No. PCT/US98/24138	**Applicant's or agent's file reference** 881.003WO1
International filing date (day/month/year) 12 November 1998 (12.11.98)	**Priority date** (day/month/year) 12 November 1997 (12.11.97)
Applicant CHIANG, Vincent, Lee, C. et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

11 June 1999 (11.06.99)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Olivia RANAIVOJAONA

Telephone No.: (41-22) 338.83.38

17

P[redacted]ENT COOPERATION TREA[redacted]

PCT

REC'D 14 MAR 2000
WIPO PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
881.003WO1		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US98/24138	12/11/1998	12/11/1997
International Patent Classification (IPC) or national classification and IPC C12N15/82		
Applicant BOARD OF CONTROL OF MICHIGAN TECHNOL. ..... et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 8 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☒ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☒ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☐ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
11/06/1999	09.03.00
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Alt. G Telephone No. +49 89 2399 8545

Form PCT/IPEA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/US98/24138

I. Basis of the report

1. This report has been drawn on the basis of (*substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

Description, pages:

1-38 as originally filed

Claims, No.:

1-53 as originally filed

Drawings, sheets:

1/14-14/14 as originally filed

2. The amendments have resulted in the cancellation of:

☐ the description, pages:
☐ the claims, Nos.:
☐ the drawings, sheets:

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

4. Additional observations, if necessary:

IV. Lack of unity of invention

1. In response to the invitation to restrict or pay additional fees the applicant has:

☐ restricted the claims.

☐ paid additional fees.

☐ paid additional fees under protest.

☒ neither restricted nor paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:

   see separate sheet

4. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this report:

   ☐ all parts.

   ☒ the parts relating to claims Nos. 1-28, 33-44, 50-53.

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	3,17,34, 33-35,37,50,52
	No:	Claims	1,2,4-16, 18-28, 36, 38-44, 51, 53
Inventive step (IS)	Yes:	Claims	
	No:	Claims	1-28, 33-44, 50-53
Industrial applicability (IA)	Yes:	Claims	1-28, 33-44, 50-53
	No:	Claims	

2. Citations and explanations

   see separate sheet

VI. Certain documents cited

1. Certain published documents (Rule 70.10)

   and / or

2. Non-written disclosures (Rule 70.9)

   see separate sheet

Re Item I

Basis of the opinion

1. It is noted that the numbering of the claims taken as a basis for the present opinion is the one presented with the set of claims substituted under Rule 26 PCT.

Re Item IV

Lack of unity of invention

2. The IPEA considers that the present claims do not relate to one invention or a group of inventions so linked as to form a single general inventive concept as required by Rule 13.1 PCT. The reasoning is as follows:
Currently, the inventive concept linking all claims can be considered as "methods for altering the growth characteristics of a plant by incorporating into the genome of a plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase or regulatory parts thereof".

This concept is however known from Kajita, S. et al., Plant Cell Physiology, vol. 37, no. 7 (1996), pages 957-965. The document discloses that the introduction of 4-coumarate:coenzyme A ligase (4CL) chimeric sense and antisense genes into tobacco caused the reduction of 4CL activity. The observed effects were that the cell walls of the xylem tissue in stems were brown, that the molecular structure of lignin in the coloured cell walls was different from that of control plants and that the lignin content was reduced.

Thus, since the above defined inventive concept is not novel, the application is considered as being directed to nine different inventions which are not linked by corresponding special technical features. The specific features (in bold letters) are regarded to be:

Invention 1:
Claims 1-17: Incorporation into the genome of a plant a nucleotide sequence encoding 4-coumarate-Co-enzyme A ligase for **altering the growth characteristics**.

Invention 2:
Claims 18-28: **Genetically down regulating** the enzyme 4-coumarate Co-enzyme A ligase for altering the characteristic of a plant, the characteristic selected from the group of **accelerated growth, reduced lignin content, altered lignin structure, increased disease resistance and increased cellulose content.**

Invention 3:
Claims 29-31, 45, 46, 48: A DNA molecule comprising a DNA segment comprising a **transcriptional regulatory region** of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and **directing expression to the xylem.**

Invention 4:
Claims 29, 30, 32, 45, 47: DNA molecule comprising a DNA segment comprising a **transcriptional regulatory region** of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and **directing expression to epidermal tissue.**

Invention 5:
Claims 33-38, 49: Introducing an **expression cassette** comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase **for imparting disease resistance.**

Invention 6:
Claims 39, 40: Introducing an **expression cassette** comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase **for altering the lignin content.**

Invention 7:
Claims 41, 42: Introducing an **expression cassette** comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase **for altering the cellulose content.**

Invention 8:

Claims 43, 44: Introducing an **expression cassette** comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase **for altering the lignin structure.**

Invention 9:
Claims 50-53: Incorporating into the genome of the plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase **for enhancing root growth.**

Although there are formally nine inventions encompassed in the present application, the IPEA considers that the inventions reflected by claims 1-28, 33-44 and 50-53 on the one hand and the inventions reflected by claims 29-32 and 45-48 on the other hand could be examined together without effort justifying the payment of eight additional fees. Therefore, the Applicant was invited to pay **one** additional examination fee. The Applicant has not reacted to this invitation. Therefore, only the claims 1-28, 33-44 and 50-53 are examined in this IPER.

Re Item V

Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

3. KAJITA, S. ET AL.: 'Alterations in the biosynthesis of lignin in transgenic plants with chimeric gens for 4-coumarate:coenzyme A ligase' PLANT CELL PHYSIOLOGY, vol. 37, no. 7, 1996, pages 957-965 (hereinafter referred to as **D1**) discloses that the introduction of 4-coumarate:coenzyme A ligase (hereinafter referred to as "4CL") chimeric sense and antisense genes (see page 958, second column, lines 3-12) into tobacco caused reduction of 4CL activity (see for example page 963, first column, first paragraph). The observed effects were that cell walls of the xylem tissue in stems were brown (page 960, first column, third paragraph), that the molecular structure of lignin in the coloured cell walls was different from that of control plants (page 963, second paragraph, last sentence) and that the lignin content was reduced (page 962, second column).
Consequently, D1 is regarded as novelty destroying for the subject-matter of claims 1,2, 4-16, 18-28, 36, 38-44, 51 and 53 (Article 33(2) PCT).

4. DOUGLAS, C.J. ET AL.: 'Exonic sequences are required for elicitor and light activation of a plant defense gene, but promoter sequences are sufficient for tissue specific expression' THE EMBO JOURNAL, vol. 10, no. 7, July 1991, pages 1767-1775 (hereinafter referred to as **D2**) discloses tobacco plants transformed with a complete parsley 4CL-1 genomic clone (page 1770, second column, first paragraph). Expression of the gene was detected (page 1770, last sentence). Consequently, the subject-matter of claims 10, 12, 13, 15, 16, 36, 38, 40, 42, 44, 51 and 53 (Article 33(2) PCT).

5. It is noted that some of the effects on plants expressing 4-CL gene are not mentioned in either of D1 or D2 (for example claim 36, "imparting disease resistance" or claim 51, "enhanced root growth"). However, the fact that an effect is not stated in a document does not necessarily render the subject-matter of a claim stating this effect novel over that document. Moreover, it is stated in the present application that the additional expression of the 4-CL gene automatically provide the plants with the mentioned effects.

6. It follows from the above evaluation that the subject-matter of claims 3, 17, 33-35, 37, 50 and 52 is novel.

7. It appears that none of the features of the above mentioned novel claims when combined with features of any other claim would be appropriate to render the subject-matter of the novel claims inventive.

Claim 3: expression of heterologous 4-CL genes is known from D2 (parsley gene in tobacco) or LEE, D. ET AL.: 'The Arabidopsis thaliana 4-coumarate:CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA' PLANT MOLECULAR BIOLOGY, vol. 28, 1995, pages 871-884 (hereinafter referred to as **D3**).
Claim 17: transformation of any plant, i.e. plants that will become trees, is common general knowledge
Claims 33-35: The role of 4-CL in disease resistance is for example known from UHLMANN,A. AND EBEL, J.: 'Molecular cloning and expression of 4-coumarate:coenzyme A ligase, an enzyme involved in the resistance response of soybean (Glycine max L.) against pathogen attack' PLANT PHYSIOLOGY

(hereinafter referred to as **D4**).
Claim 37: It is common general knowledge that normally seeds are produced from transgenic plants.
Claims 50 and 52: The effect of 4-CL on root growth can for example be taken from D3, page 876, second column, last full sentence; page 877, first column, last sentence of second paragraph.

Consequently, at present, the subject-matter of claims 3, 17, 33-35, 37, 50 and 52 is not regarded to involve an inventive step (Article 33(3) PCT).

Re Item VI
Certain documents cited
WO-A-9811205

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	FOR FURTHER ACTION	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
881.003WO1		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/US 98/24138	12/11/1998	12/11/1997
Applicant		
BOARD OF CONTROL OF MICHIGAN TECHNOL. ..... et al.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 8 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language,** the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   [X] contained in the international application in written form.

   [ ] filed together with the international application in computer readable form.

   [ ] furnished subsequently to this Authority in written form.

   [X] furnished subsequently to this Authority in computer readble form.

   [X] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   [X] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1

   [X] as suggested by the applicant.

   [ ] because the applicant failed to suggest a figure.

   [ ] because this figure better characterizes the invention.

   [ ] None of the figures.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 98/24138

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-50 (1-17)

Method for altering the growth characteristics of a plant by incorporating into the genome a DNA molecule comprising a nucleotide sequence encoding 4-coumarate-Co-enzyme A ligase and corresponding plants.

2. Claims: 18-25

Method for altering the characterisitic of a plant, the characteristic selected from the group of accelerated growth, reduced lignin content, altered lignin structure, increased disease resistance and increased cellulose content, by genetically down-regulating the enzyme 4-coumarate Co-enzyme A ligase and corresponding plants

3. Claims: 26-28, 42, 43, 45

A DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and directing expression to the xylem.

4. Claims: 26, 27, 29, 42, 44

DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and directing expression to epidermal tissue

5. Claims: 30-35

Method of imparting disease resistance to a plant tissue by introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase and corresponding plants and seeds

6. Claims: 36, 37

Method for altering the lignin content in a plant by introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase and corresponding plants

7. Claims: 38, 39

Method for altering the cellulose content in a plant by introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase and corresponding plants

8. Claims: 40, 41

Method for altering the lignin structure in a plant by introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase and corresponding plants

9. Claims: 47-50

Method for enhancing the root growth of a plant by incorporating into the genome of the plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase and corresponding plants

The ISA considers that the present claims do not relate to one invention or a group of inventions so linked as to form a single general inventive concept as required by Rule 13.1 PCT. The reasoning is as follows:
Currently, the inventive concept linking all claims can be considered as methods for altering the growth characteristics of a plant by incorporating into the genome of a plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase or regulatory parts thereof.

This concept is however known from Kajita, S. et al., Plant Cell Physiology, vol. 37, no. 7 (1996), pages 957-965. The document discloses that the introduction of 4-coumarate:coenzyme A ligase (4CL) chimeric sense and antisense genes into tobacco caused the reduction of 4CL acitivty. The observed effects were that the cell walls of the xylem tissue in stems were brown, that the molecular structure of lignin in the colored cell walls was different from that of control plants and that the lignin content was reduced.

Thus, since the above defined inventive concept is not novel, the application is considered as being directed to nine different inventions which are not linked by corresponding special technical features. The specific features are:
1. Claims 1-17: Incorporation into the genome a nucleotide sequence encoding 4-coumarate-Co-enzyme A ligase for altering the growth characteristics.

2. Claims 18-25: Genetically down regulating the enzyme 4-coumarate Co-enzyme A ligase for altering the characteristic of a plant, the characteristic selected from the group of accelarated growth, reduced lignin content, altered lignin structure, increased disease

resistance and increased cellulose content.

3. Claims 26-28, 42, 43, 45: A DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and directing expression to the xylem.

4. Claims 26, 27, 29, 42, 44: DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and directing expression to epidermal tissue.

5. Clais 30-35, 46: Introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase for imparting disease resistance.

6. Claims 36, 37: Introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase for altering the lignin content.

7. Claims 38, 39: Introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase for latering the cellulose content.

8. Claims 40, 41: Introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase for altering the lignin structure.

9. Claims 47-50: Incorporating into the genome of the plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase for enhancing root growth.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/[redacted]98/24138

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N15/11 C12N15/63 A01H1/00 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 A01H C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DOUGLAS, C.J. ET AL.: "Exonic sequences are required for elicitor and light activation of a plant defense gene, but promoter sequences are sufficient for tissue specific expression" THE EMBO JOURNAL, vol. 10, no. 7, July 1991, pages 1767-1775, XP002100277 see page 1770, sec. column; page 1771,first column,lines 7-13 and sec. column, lines 11-14;page 1773, sec. column, lines 6-8 and last paragraph contind. on page 1774, lines 1-5; page 1774, first column, lines 24-26, sec. column, first paragraph;	1,3,4,6, 7,10, 12-15, 26, 28-31, 33,42-44
A	page 1771, first column, lines 11-12	47-50

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 April 1999	05. 05. 99

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Alt, G

3

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KAJITA, S. ET AL.: "Alterations in the biosynthesis of lignin in transgenic plants with chimeric gens for 4-coumarate:coenzyme A ligase" PLANT CELL PHYSIOLOGY, vol. 37, no. 7, 1996, pages 957-965, XP002065207 page 958, second column, lines 1-12; Figure 2;page 960, sec. column, sec. paragraph;page 961, sec. column; 962, sec.column; page 963, first column, first paragraph	1,2, 4-16, 18-26, 36-42
A	ZHANG, X.-H. AND CHIANG, V.L.: "Molecular cloning of 4-coumarate:coenzym A Ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood" PLANT PHYSIOLOGY, vol. 113, January 1997, page 65-74 XP002100278 see page 70, first column, lines 33-36	28, 36-41,43
A	UHLMANN,A. AND EBEL, J.: "Molecular cloning and expression of 4-coumarate:coenzyme A ligase, an enzyme involved in the resistance response of soybean (Glycine max L.) against pathogen attack" PLANT PHYSIOLOGY, XP002100279 see the abstract	1
A	LEE, D. ET AL.: "The Arabidopsis thaliana 4-coumarate:CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA" PLANT MOLECULAR BIOLOGY, vol. 28, 1995, pages 871-884, XP002100280 see page 876, second column, lines 16-20	47-50
P,X	WO 98 11205 A (GENESIS RESEARCH & DEVELOPMENT CORPORATION LIMITED) 19 March 1998 see Example 4	3-7, 10-14, 18-20, 22,24, 25,36,37

-/--

INTERNATIONAL SEARCH REPORT

International Application No
PCT/[redacted]98/24138

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	HU, W.J. ET AL.: "Compartmentalized expression of two structuraly and functionally distinct 4-coumarate:CoA ligase genes in aspen (Populus tremoides)" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 95, April 1998, pages 5407-5412, XP002100281 see the whole document -----	1

3

INTERNATIONAL SEARCH REPORT

Information [redacted]atent family members

International Application No
PCT/[redacted]98/24138

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9811205 A	19-03-1998	US 5850020 A	15-12-1998
		AU 4403697 A	02-04-1998

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/00

A2

(11) International Publication Number: WO 99/24561

(43) International Publication Date: 20 May 1999 (20.05.99)

(21) International Application Number: PCT/US98/24138

(22) International Filing Date: 12 November 1998 (12.11.98)

(30) Priority Data:
08/969,046 12 November 1997 (12.11.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/969,046 (CIP)
Filed on 12 November 1997 (12.11.97)

(71) Applicant *(for all designated States except US)*: BOARD OF CONTROL OF MICHIGAN TECHNOLOGICAL UNIVERSITY [US/US]; 1400 Townsend Drive, Houghton, MI 49931 (US).

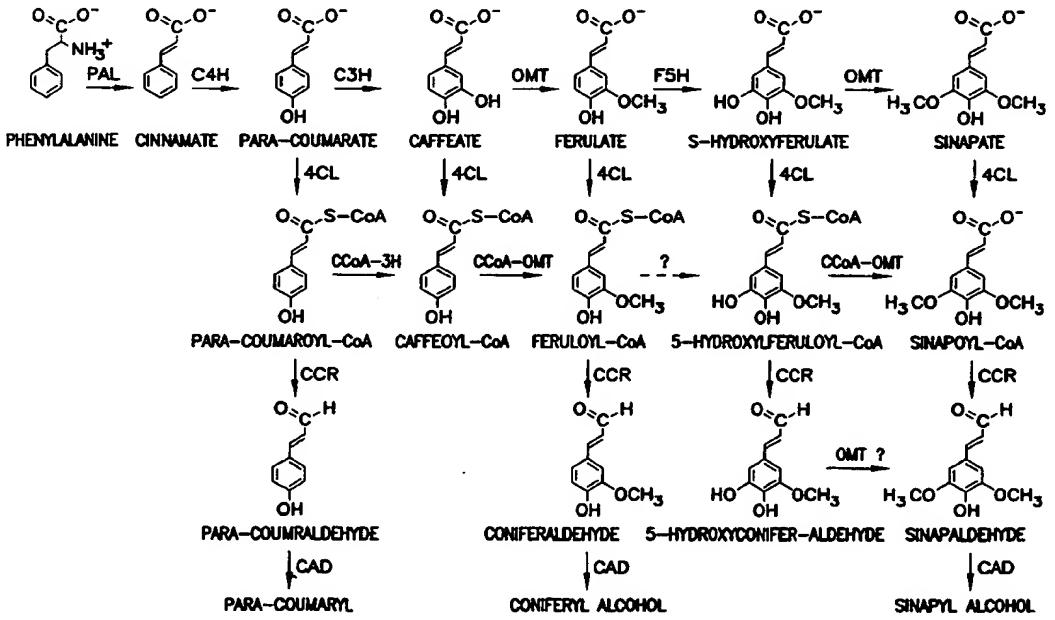
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: CHIANG, Vincent, Lee, C. [US/US]; 1104 Birch Street, Hancock, MI 49930 (US). TSAI, Chung-Jui [-/US]; 1510 Anthony Street #34, Hancock, MI 49930 (US). HU, Wen-Jing [-/US]; 8299 Cambridge Drive #1003, Houston, TX 77054 (US).

(74) Agent: VIKSNINS, Ann, S.; Schwegman, Lundberg, Woessner & Kluth, P.O. Box 2938, Minneapolis, MN 55402 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GENETIC ENGINEERING OF LIGNIN BIOSYNTHESIS IN PLANTS



(57) Abstract

The invention pertains to altering a lignin pathway 4-coumarate Co-enzyme A ligase (4CL) in plants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GENETIC ENGINEERING OF LIGNIN BIOSYNTHESIS IN PLANTS

Cross-Reference to Related Applications

This is a continuation-in-part of U.S. application Serial No. 08/969,046, filed November 12, 1997, the disclosure of which is incorporated by reference herein.

Field of the Invention

The invention relates to genetically modifying plants, e.g., trees, through manipulation of the lignin biosynthesis pathway, and more particularly, to genetically modifying plants through the down regulation of 4-coumarate Co-enzyme A ligase (4CL) to achieve faster growth. Down regulation of 4CL may also achieve altered lignin content, and/or altered lignin structure, and/or altered cellulose content, and/or altered disease resistance of the trees. Moreover, promoters of the 4CL genes are useful to drive gene expression specifically in xylem tissue or specifically in epidermal tissues.

Background of the Invention

Genetic engineering of plants to conform to desired traits has shifted the emphasis in plant improvement away from the traditional breeding programs during the past decade. Although research on genetic engineering of plants has been vigorous, the progress has been slow.

The ability to make plants grow faster continues to be the top objective of many companies worldwide. The ability to genetically increase the optimal growth of plants would be a commercially significant improvement. Faster growing plants could be used by all sectors of the agriculture and forest products industries worldwide.

Lignin, a complex phenolic polymer, is a major component in cell walls of secondary xylem. In general, lignin constitutes 25% of the dry weight of the wood, making it the second most abundant organic compound on earth after cellulose. Although lignin plays an important role in plants, it usually represents an obstacle to utilizing biomass in several applications. For example, in wood

pulp production, lignin has to be removed through expensive and polluting processes in order to recover cellulose.

Thus, it is desirable to genetically engineer plants with reduced lignin content and/or altered lignin composition that can be utilized more efficiently. Plants that could be genetically engineered with a reduced amount of lignin would be commercially valuable. These genetically engineered plants would be less expensive to pulp because, in essence, part of the pulping has already been performed due to the reduced amount of lignin. Further, plants with increased cellulose content would also be commercially valuable to the pulp and paper industry.

Disease resistance in plants is also a desirable plant trait. The impact of disease resistance in plants on the economy of plant products industry worldwide is significant.

Thus, what is needed is the identification and characterization of genes useful to enhance plant growth, alter lignin content and/or structure in plants, alter cellulose content in plants, and/or provide or enhance disease resistance of plants.

Summary of the Invention

The invention provides a method to genetically alter plants through the down regulation (decrease) or inhibition of native (endogenous) 4-coumarate Co-enzyme A ligase (4CL) in that plant. Such down regulation of 4-coumarate Co-enzyme A ligase results in faster growth, and/or reduced lignin content, and/or altered lignin structure, and/or altered cellulose content, and/or altered disease resistance in the genetically altered plant. The invention also provides for genetically engineered plants, e.g., transformed or transgenic plants, which have been altered to down regulate or inhibit native 4-coumarate Co-enzyme A ligase in the plant so as to achieve faster growth, and/or reduced lignin content, and/or altered lignin structure, and/or increased cellulose content, and/or increased disease resistance. Preferred genetically altered plants include trees, e.g., angiosperms or gymnosperms, forage crops, and more preferably a forest tree, e.g., Populus. Preferred angiosperms include, but are not limited to, Populus, Acacia, Sweetgum, yellow poplar, maple and birch, including pure

lines and hybrids thereof. Preferred gymosperms include, but are not limited to, Pine, Spruce, Douglas-fir and hemlock.

The invention further provides a transgenic plant, the genome of which is augmented by a recombinant DNA molecule encoding 4-coumarate Co-enzyme A ligase, or a recombinant DNA molecule comprising an antisense 4-coumarate Co-enzyme A ligase gene, or a fragment thereof. The recombinant DNA molecule is expressed so as to down regulate, decrease or inhibit lignin pathway 4-coumarate Co-enzyme A ligase.

The invention also provides an isolated and purified DNA molecule comprising a DNA segment comprising a transcriptional regulatory control region of a 4-coumarate Co-enzyme A ligase gene. Preferably, the transcriptional regulatory region comprises a promoter. Tissue specific promoters of a 4-coumarate Co-enzyme A ligase gene can be used to manipulate gene expression in target tissue such as xylem and epidermal tissues, as described hereinbelow. Preferably, the promoter is derived from aspen DNA. Therefore, the invention also provides an expression cassette comprising a transcriptional regulatory region of a 4-coumarate co-enzyme A ligase gene, a method of using the region to express a preselected DNA segment in a tissue-specific manner in plant cells, and a transgenic plant comprising the expression cassette.

Also provided is a method to alter, e.g., enhance, plant growth. The method comprises introducing an expression cassette into cells of a plant, e.g., the cells of a tree, so as to yield genetically altered plant cells. The expression cassette comprises a recombinant DNA molecule, segment, or sequence, comprising a 4-coumarate Co-enzyme A ligase gene, or a fragment thereof. Preferably, the 4-coumarate Co-enzyme A ligase gene, or fragment thereof, is in antisense orientation. The 4-coumarate Co-enzyme A ligase gene may be a homologous or heterologous 4-coumarate Co-enzyme A ligase gene. The transformed plant cells are regenerated to provide a genetically altered, e.g., transgenic, plant. The recombinant DNA is expressed in the cells of the regenerated, genetically altered plant in an amount that confers enhanced or accelerated growth to the regenerated, genetically altered plant relative to the

corresponding non-genetically altered plant. Preferably, the genetically altered plant is a tree. It is preferred that a genetically altered tree of the invention has an increase in height, leaf size, diameter and/or average internode length relative to the corresponding non-genetically altered tree.

Hence, the invention also provides for a genetically altered plant, the genome of which is augmented by a recombinant DNA molecule encoding 4-coumarate Co-enzyme A ligase, or a recombinant DNA molecule comprising an antisense 4-coumarate Co-enzyme A ligase gene, or fragment thereof, which plant has altered growth characteristics relative to the corresponding non-genetically altered plant.

Further provided is a method to genetically alter plants so as to change or alter their lignin structure. The method comprises introducing an expression cassette into cells of a plant, e.g., a tree, so as to yield genetically altered plant cells. The expression cassette preferably comprises an antisense recombinant DNA molecule, segment or sequence comprising a 4-coumarate Co-enzyme A ligase gene, or a fragment thereof. The transformed plant cells are regenerated to provide a regenerated, genetically altered plant. The recombinant DNA is expressed in the cells of the regenerated, genetically altered plant in an amount that alters the lignin structure in the cells of the plant relative to the corresponding non-genetically altered plant.

Also provided is a method for altering the lignin content in a plant. The method comprises introducing an expression cassette comprising a recombinant DNA molecule comprising a 4-coumarate Co-enzyme A ligase gene operably linked to a promoter functional in a plant cell into the cells of a plant. The plant cells are regenerated so as to yield a genetically altered plant. The recombinant DNA molecule is expressed in the cells of the regenerated plant in an amount effective to alter the lignin content in the plant cells. Preferably, the lignin content is reduced. Also preferably, the lignin content is reduced in a tissue-specific manner. In particular, a reduction in lignin content in forage crops is useful as the digestability of these crops by ruminants is increased. Also preferably, the 4-coumarate Co-enzyme A ligase gene is in an antisense orientation relative to the promoter.

Further provided is a genetically altered, e.g., transgenic, plant having an altered lignin content in the plant cells. The plant comprises a recombinant DNA molecule comprising a nucleotide sequence encoding a plant 4-coumarate Co-enzyme A ligase operably linked to a promoter so that the recombinant DNA molecule is expressed in an amount effective to alter the lignin content of the plant.

Yet another embodiment of the invention is a method to alter, e.g., increase, the cellulose content in plants. The method comprises introducing an expression cassette into cells of a plant, e.g., a tree, so as to yield genetically altered plant cells. The expression cassette preferably comprises an antisense recombinant DNA molecule, segment or sequence comprising a 4-coumarate Co-enzyme A ligase gene, or a fragment thereof, operably linked to a promoter functional in a plant cell. The transformed plant cells are regenerated to provide a regenerated, genetically altered plant. The recombinant DNA is expressed in the cells of the regenerated, genetically altered plant in an amount that alters the cellulose content in plant. Thus, the invention further provides a genetically altered, e.g., transgenic, plant having an altered cellulose content.

The invention also provides a method to genetically alter plants to increase their disease resistance, e.g., to fungal pathogens. The method comprises introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase operably linked to a promoter functional in a plant cell into cells of a plant. The transformed plant cells are regenerated to provide a genetically altered plant. The recombinant DNA molecule is expressed in the cells of the regenerated, genetically altered plant in an amount effective to render the plant resistant to disease. Preferably, the recombinant DNA molecule is expressed in amount that decreases the amount of lignin in the plant and/or increases the amount of phenolic compounds which are toxic to fungal pathogens. Hence, the invention also provides a transgenic plant, which is substantially resistant to disease. The plant comprises a native 4-coumarate Co-enzyme A ligase gene, and a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase operably linked to a promoter functional in a

plant wherein the recombinant DNA molecule is expressed in an amount effective to confer resistance to the transgenic plant.

Other features and advantages of the invention will become apparent to those of ordinary skill in the art upon review of the following drawings, detailed description and claims.

Brief Description of the Drawings

Fig. 1 is a schematic of a phenylpropanoid pathway;

Fig. 2 is a diagram of Agrobacterium T-DNA gene construct pA4CL1;

Fig. 3 is a restriction map of genomic clone Pt4CL1g-4;

Fig. 4 is a restriction map of genomic clone Pt4CL2g-11;

Fig. 5 is a restriction map of subcloned pT4CL1 gene promoter p7Z-4XS;

Fig. 6 is a restriction map of subcloned pT4CL2 gene promoter pSK-11HE;

Fig. 7 is an Agrobacterium T-DNA construct of Pt4CL1 promoter and GUS fusion gene Pt4CL1p-GUS; and

Fig. 8 is an Agrobacterium T-DNA construct of Pt4CL2 promoter and GUS fusion gene, Pt4CL2p-GUS.

Fig. 9 shows biosynthetic pathways to guaiacyl (coniferyl alcohol **9a**) and syringyl (sinapyl alcohol **9b**) monolignols for the formation of guaiacyl-syringyl lignin in wood angiosperms. Enzymes are indicated for each reaction step. C4H, cinnamic acid 4-hydroxylase; C3H, 4-coumaric acid 3-hydroxylase; COMT, caffeic acid *O*-methyltransferase; F5H, ferulic acid 5-hydroxylase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase. Aspen 4CL (Pt4CL1) converts 4-coumaric **2**, caffeic **3,** ferulic **4**, 5-hydroxyferulic **5**, and sinapic **6** acids into their corresponding thioesters for the formation of feruloyl-CoA **7a** and sinapoyl-CoA **7b**, leading to **9a** and **9b**, respectively.

Fig. 10. The effects of down-regulation of Pt4CL1 expression on Pt4CL1 activity and lignin accumulation in transgenic aspen. (A) Northern blot analysis of Pt4CL1 transcript levels in control (lane C) and transgenic aspen (3, 4, 5, 6, 8, and 9). Each lane contained 20 μg of total RNA extracted from developing xylem and the blot was hybridized (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY (1989)) with

Pt4CL1 cDNA. (B) Pt4CL1 enzyme activities in developing xylem tissues. Crude protein (40 μg) extracted from xylem tissue was assayed spectrophotometically for Pt4CL1 activities with various hydroxylated cinnamic acids (Ranjeva et al., 1976). Error bars represent SD values of three replicates. (C) Levels of lignin reduction in woody stem of transgenic lines as compared to the control, based on the lignin contents presented in Table 7. (D and E) Fluorescence microscopy showing transverse sections of the 20th internode from control (D) and transgenic line 6 (E). Lignin autofluorescence was visualized following UV-excitation at 365 nm.

Fig. 11 depicts regions of the HSQC spectra (NMR experiments were performed at 360 MHZ on a Bruker DRX-360 using a narrow bore probe with inverse coil geometry (proton coils closest to the sample) and with gradients. Experiments used were standard Bruker implementations of gradient-selected inverse (^{1}H-detected) HSQC (Palmer et al., J. Magn. Reson. Ser. A, 111, 70 (1991)), HSQC-TOCSY (Braunschweiler et al., J. Magn. Reson., 53, 521 (1983)), and HMQC (Ruiz-Cabello et al., J. Magn. Reson., 100, 282 (1992)) along with the standard 1D ^{13}C (proton-decoupled) and ^{1}H NMR experiments. TOSCY experiments used a 100 ms spin lock period; HMBC used either an 80 or a 100 ms long-range coupling delay.) of isolated milled wood lignins from (A) control and (B) transgenic line 6. Structure assignments (Ralph et al., 1997) reveal the existence of some major structural units in both samples that are common to angiosperm lignin. The *erytho*-(δ_{Ca}/δ_{Ha}:75.4/6.05) and *threo*-(δ_{Ca}/δ_{Ha}:76.6/6.08) isomers of β-aryl ethers 10 are indicated. 5-5-Homo-coupling of coniferyl alcohol 9a involved in dibenzodioxocins 13 (δ_{Ca}/δ_{Ha}:85.3/4.94) (Ralph et al., 1997) was not detected in either sample. Yellow contours are from intense methoxyl signals and light green contours form xylan residues. Other components (gray contours) in both lignin samples, not relevant or not identified, are commonly seen in many other angiosperm lignin preparations.

Fig. 12 shows enhanced growth in transgenic aspen. (A) 10-Week-old plants of control and four transgenic aspen grown in a greenhouse (ruler = 25 cm). (B) Control and transgenic leaves from the 10th internodes. (C to F) SE images of stem transverse sections of control [C (bar = 50 μm) and E (bar = 10

μm)] and transgenic line 6 [D (bar = 50 μm) and F (bar = 10 μm)]. (G) 2-week-old *ex vitro* rooted stem cuttings from control and transgenic aspen lines 5 and 6. Two cuttings from each line are shown. (H) Leaf upper epidermal cell area. Values represent the mean of at least 100 determinations per leaf. Sample SD was 15 to 20% of the mean for all determinations.

Before one embodiment of the invention is explained in detail, it is to be understood that the invention is not limited in its application to the details set forth in the following description of the preferred embodiment. The invention is capable of other embodiments and of being practiced or being carried out in various ways. Also, it is to be understood that the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting.

Detailed Description of the Invention

The invention pertains to genetically down regulating a lignin pathway 4-coumarate Co-enzyme A ligase (4CL) in a plant. Plants which have been genetically transformed to down regulate 4CL will hereafter be called transgenic plants. Such down regulation can result in faster growing plants. Such down regulation can also result in a reduction in the lignin content of the plants and/or altered lignin structure. Such down regulation can further result in increased cellulose content. Such down regulation may also result in increased disease resistance. Further, by using a specific 4CL promoter, targeted tissue-specific gene expression can be achieved in either the xylem or the epidermal tissues of the plant.

A. 4CL

Lignin is synthesized by the oxidative coupling of three monolignols (coumaryl, coniferyl and sinapyl alcohols) formed via the phenylpropanoid pathway as shown in Fig. 1. Reactions in the phenylpropanoid pathway include the deamination of phenylalanine to cinnamic acid followed by hydroxylations, methylations and activation of substituted cinnamic acids to coenzyme A (CoA) esters. The CoA esters are then reduced to form monolignols which are secreted from cells to form lignin.

The products of the phenylpropanoid pathway are not only required for the synthesis of lignin but also required for the synthesis of a wide range of aromatic compounds including flavonoids, phytoalexins, stilbenes and suberin.

In the phenylpropanoid pathway, 4CL activates a number of cinnamic acid derivatives, including 4-coumaric acid, caffeic acid, ferulic acid, 5-hydroxyferulic acid and sinapic acid. The resulting products, CoA esters, serve as substrates for entry into various branch pathways, such as lignin, flavonoids, phytoalexins, stilbenes and suberin. The esterification reactions catalyzed by 4CL require high energy and the reactions are not likely to occur without 4CL. 4CL is important in making a continuous flow of the lignin biosynthesis pathway. 4CL is also important because it is located at the branching points of the phenylpropanoid metabolism. 4CL is suggested to play a pivotal role in regulating carbon flow into specific branch pathways of the phenylpropanoid metabolism in response to stages of development and environmental stress.

The basic properties of 4CL are quite uniform. 4CL depends on ATP as a cosubstrate and requires Mg^{2+} as a cofactor. The optimal pH for 4CL ranges from pH 7.0 to 8.5 and the molecular weight of 4CL isoforms from various plant species ranges from 40 kD to 75 kD. Most 4CLs have high affinity for substituted cinnamic acids. 4CL has the highest activity with 4-coumaric acid.

4CL cDNA sequences have been reported for parsley, potato, soybean, rice, loblolly pine, *Arabidopsis, Lithosperum, Vanilla* and tobacco. 4CL genes have been isolated and sequenced for parsley, rice, potato and loblolly pine. The analysis of 4CL cDNAs and genes indicates that 4CL is encoded by multiple/divergent genes in rice, soybean, and *Lithosperum*, very similar genes in parsley, potato, tobacco, and loblolly pine, and a single gene in *Arabidopsis*. Two similar 4CL cDNAs in parsley, potato and tobacco have been shown to be expressed at similar level in response to environmental stress and during different developmental stages. Two distinct 4CL cDNAs in soybean and Lithosperum have shown different expression levels when pathogens or chemicals were applied to the cell cultures. It appears that the expression of the 4CL genes is developmentally regulated and inducible by many environmental

stresses at the transcription level. 4CL promoters have been isolated and sequenced for parsley, rice and potato.

Alignment of deduced amino acid sequences of cloned plant 4CL sequences reveals two highly conserved regions. The first conserved region (LPYSSGTTGLPK; SEQ ID NO:7), proposed to designate a putative AMP-binding region, consists of a serine/theronine/glycine-rich domain followed by a proline/lysine glycine triplet. The second conserved region (GEICIRG; SEQ ID NO:8) contains one common Cys residue. The amino acid sequences of 4CL from plants contain a total of five conserved Cys residues.

The description of the invention hereafter refers to the tree species aspen, and in particular quaking aspen (*Populus tremuloides* Michx), when necessary for the sake of example. However, it should be noted that the invention is not limited to genetic transformation of trees such as aspen. The method of the present invention is capable of being practiced for other plant species, including for example, other angiosperm, and other gymnosperm forest plants species, legumes, grasses, other forage crops and the like.

Preferably, the 4CL down regulation is accomplished through transformation with a homologous 4CL sequence in an antisense orientation. However, it should be noted that a heterologous antisense 4CL sequence could be utilized and incorporated into a plant species to down regulate 4CL if the heterologous 4CL gene sequence has a high nucleotide sequence homology or identity of at least about 70%, more preferably at least about 80%, and more preferably at least about 90%, to the endogenous (native) 4CL gene sequence of that plant species, e.g., a tree species.

In addition, plants transformed with a sense 4CL sequence may also cause a sequence homology-based cosuppression of the expression of the transgene and endogenous 4CL gene, thereby achieving down regulation of 4CL in these plants.

B. Isolation of 4CL cDNAs

The present invention utilizes a homologous 4CL sequence to genetically alter plants. The example described below utilizes a cDNA clone of the quaking aspen 4CL gene to genetically alter quaking aspen.

Two 4CL cDNAs, Pt4CL1 and Pt4CL2, have been isolated from quaking aspen. Pt4CL1 cDNA is lignin pathway-specific and is different from Pt4CL2 cDNA, which is involved in flavonoid synthesis. It should be noted that the methods described below are set forth as an example and should not be considered limiting. The sequences of these 4CL cDNA clones are available from Genbank, Accession Nos. AF041049 and AF041050.

Pt4CL1 and Pt4CL2 genomic clones including their 5' -end regulatory promoter sequences were also isolated. The promotor of Pt4CL 1 (Pt4CL1p) directs xylem tissue-specific gene expression in a plant, whereas the promoter of Pt4CL2 (Pt4CL2p) drives the expression of genes specifically in epidermal tissues of stem and leaf of a plant. These tissue specific promoters will be discussed in more length below.

Young leaves and shoot tips are collected from greenhouse-grown quaking aspen (*Populus tremuloides* Michx). Differentiating xylem is collected from three to four year old quaking aspen. The bark is peeled from the tree exposing the developing secondary xylem on the woody stem. Developing secondary xylem is scraped from the stem and bark with a razor blade and immediately frozen in liquid nitrogen until use.

Total RNA is isolated from the young leaves, shoot tips, and xylem following the method of Bugos et al., Biotechniques 19(5):734-737 (1995). Poly(A)$^+$ RNA is purified from total RNA using Poly(A)$^+$ mRNA Isolation Kit from Tel-test B, Inc. A unidirectional Lambda gt22 expression cDNA library was constructed from the xylem mRNA using Superscript λ System from Life Technologies, Inc. and Gigapack Packaging Extracts from Stratagene. The Pt4CL1 cDNA was obtained by screening the cDNA library with a ^{32}P-labeled parsley 4CL cDNA probe. The parsley 4CL cDNA (pc4CL2) has Genbank Accession No. X13325.

The Pt4CL2 cDNA was obtained by RT-PCR. The reverse transcription of total RNA isolated form shoot tips was carried out using the Superscript II reverse transcriptase from Life Technologies. Two sense primers (R1S, 5'-TTGGATCCGGIACIACIGGIYTICCIAARGG-3'; SEQ ID NO:9 and H1S, 5'-TTGGATCCGTIGCICARCARGTIGAYGG-3'; SEQ ID NO:10) were designed

around the first consensus AMP-binding region of 4CL that was previously discussed. One antisense primer (R2A, 5′-ATGTCGACCICKDATRCADATYTCICC-3′; SEQ ID NO:11) was designed based on the sequence of the putative catalytic motif GEICIRG (SEQ ID NO:12). One fifth of the reverse transcription reaction (4 μl) is used as the template in a 50 μl PCR reaction containing 1X reaction buffer, 200 μM each deoxyribonucleotide triphosphate, 2 μM each R1S and oligo-dT (20 mer) primers, and 2.5 units of *Taq* DNA polymerase. The PCR reaction mixture was denatured at 94°C for 5 minutes followed by 30 cycles of 94°C/45 seconds, 50°C/1 minute, 72°C/90 seconds and is ended with a 5 minute extension at 72°C. 2 μl of the PCR amplification products were used for a second run PCR re-amplification using primers H1S and R2A. A 0.6 kb PCR fragment was cloned using the TA Cloning Kit from Invitrogen and used as a probe to screen an aspen genomic library to obtain the Pt4CL2 genomic clone. Two primers (2A, 5′-TCTGTCTAGATGATGTCGTGGCCACGG-3′; SEQ ID NO:13 and 2B, 5′-TTAGATCTCTAGGACATGGTGGTGGC-3′; SEQ ID NO:14) were designed based on the genomic sequence of Pt4CL2 around the deduced transcription start site and the stop codon. These primers were used to clone Pt4CL2 cDNA by RT-PCR, as described above using total RNA isolated from shoot tips.

The DNA sequences of Pt4CL1 and Pt4CL2 cDNA were determined using ▵ *Taq* Cycle Sequencing kit from Amersham.

The Pt4CL1 cDNA has an open reading frame of 1605 bp which encodes a polypeptide of 535 amino acid residues with a predicted molecular weight of 58.498 kd and pI of 5.9. The nucleotide sequence of the aspen 4CL cDNA clone Pt4CL1 is set forth as SEQ ID NO:1. The deduced amino acid sequence for the aspen 4CL1 protein is set forth as SEQ ID NO:2.

The Pt4CL2 cDNA has an open reading frame of 1710 bp which encodes a polypeptide of 570 amino acid residues with a predicted molecular weight of 61.8 kd and pI of 5.1. The nucleotide sequence of the aspen 4CL cDNA clone Pt4CL2 is set forth as SEQ ID NO:3. The deduced amino acid sequence for the aspen 4CL2 protein is set forth as SEQ ID NO:4.

The aspen Pt4CL1 cDNA shares a 55-69% identity at the nucleotide level and 57-76% identity at the amino acid level with previously reported 4CL cDNAs and genes, whereas the Pt4CL2 cDNA shares a 60-71% identity at the nucleotide level and 58-73% at the amino acid level with other 4CL cDNAs and genes as set forth in the following table.

Table 1: Comparison of Pt4CL1 and Pt4CL2 nucleotide and predicted amino acid sequences to each other and other full length 4CL sequences.

Gene	Comparison with Pt4CL1 (%)			Comparison with Pt4CL2 (%)		
	DNA Identity	Protein Identity	Protein Similarity	DNA Identity	Protein Identity	Protein Similarity
Pt4CL1				61.3	63.4	72.7
Pt4CL2	61.3	63.4	72.7			
Le4CL1	64.5	70.7	78.1	61.8	64.6	73.4
Le4CL2	60.1	57.3	67.7	71.1	73	77.5
Nt4CL1	66	74.8	83.1	61.5	65.3	74.4
Nt4CL2	64.1	75	82.9	62.1	66.8	76
Os4CL1	59.2*	59.8	70.2	59.6*	57.7	69.5
Os4CL2	54.9	57.7	67.3	63.9	66.5	73.8
Pc4CL1	65.1	71.2	79.6	62	64.3	73.5
Pc4CL2	65	71.4	79.6	62.9	64.5	73.5
Ptd4CL1	66.6	73.7	82.2	64.5	66.6	75.8
Ptd4CL2	67	74.2	81	63.4	64.7	73.3
At4CL	63.7	69.9	78.7	62.4	61.1	70.2
Lp4CL	60.1	64	73.9	62.3	67.9	77.8
St4CL1	69.1*	74	81.4	62.2*	65.3	74.5
Vp4CL	65.2	75.5	81.6	61.5	66.5	74.1

Pt4CL1 and Pt4CLW: aspen 4CL
Le4CL1 and Le4CL2: lithospermum erythrorhizon 4CL
Nt4CL1 and Nt4CL2: tobacco 4CL
OS4CL1 and Os4CL2: rice 4CL (*.DNA sequence compared to Os4CL1 coding region only)
PC4CL1 and Pc4CL2: Parsley 4CL
Ptd4CL1 and Ptd4CL2: hybrid poplar 4CL
At4CL: Arabidopsis 4CL
Lp4CL: lobolly pine 4CL
St4CL: potato 4CL (*.DNA sequence compared to coding region only)
Vp4CL: vanilla 4CL

In a study to characterize lignification in aspen stems, it was observed that the lignin composition in the top four internodes (referred to as top internodes hereafter) was different from that in the fifth and subsequent internodes, suggesting the involvement of developmentally regulated differential expression of lignin pathway genes during the transition from primary to secondary growth in aspen stem. To investigate whether this transition regulates differential expression of 4CL gene members, 4CL genes were cloned from top and lower (6^{th}-10^{th}) internodes and secondary-developing xylem tissue of aspen stems. Nucleotide sequence analysis revealed that clones derived from lower internodes were identical to Pt4CL1, whereas clones isolated from top internodes could be divided into two groups (T1 and T2). Clones in Group T1 were found identical to Pt4CL1. Clones in group T2 shared 60-75% sequence homology with other plant 4CL genes but were distinct from Pt4CL1 cDNA and designated as Pt4CL2-600. These results together with Northern hybridization analysis suggested that Pt4CL2-600 represents a fragment of another aspen 4CL gene expressed in top internodes.

The results of sequence analysis, phylogenetic tree and genomic Southern blot analysis indicate that Pt4CL1 and Pt4CL2 cDNAs encode two distinct 4CLs that belong to two divergent gene families in aspen. The deduced amino acid sequence for the Pt4CL2 protein contains a longer N-terminal sequence than the Pt4CL1 protein but shows profound similarity in the central and C-terminal portions of protein to the Pt4CL1 protein.

Pt4CL1 and Pt4CL2 cDNAs display distinct tissue-specific expression patterns. The Pt4CL1 sequence is expressed highly in the secondary developing xylem and in the 6th to 10th internodes whereas the Pt4CL2 sequence is expressed in the shoot tip and leaves. These tissue-specific expression patterns were further investigated by fusing promoters of Pt4CL1 and Pt4CL2 genes to a GUS reporter gene. The tissue specific promoters for Pt4CL1 and Pt4CL2 are discussed in more length below.

The substrate specificity of Pt4CL1 and Pt4CL2 is also different from each other as determined using recombinant proteins produced in *E. coli.*

Pt4CL1 utilized 4-coumaric acid, caffeic acid, ferulic acid and 5-hydroxyferulic acid as substrates. Pt4CL2 showed activity for 4-coumaric acid, caffeic acid and ferulic acid but not to 5-hydroxyferulic acid.

Specifically, Pt4CL1 and Pt4CL2 were used to construct expression vectors for *E. coli* expression. The substrate specificity of Pt4CL1 and Pt4CL2 was tested using fusion proteins produced in *E. coli*. Two plasmids, pQE/4CL1 and pQE/4CL2, were constructed in which the coding regions of Pt4CL1 and Pt4CL2, respectively, were fused to N-terminal His tags in expression plasmids pQE-31 and pQE-32 (QIAGEN, Chatsworth, CA). The recombinant proteins of Pt4CL1 and Pt4CL2 produced by *E. coli* were approximately 60 kD and 63 kD, respectively.

The two recombinant proteins were tested for their activity in utilizing cinnamic acid derivatives. Pt4CL1 recombinant protein showed 100, 51, 72, 19 and 0% relative activity to 4-coumaric acid, caffeic acid, ferulic acid, 5-hydroxyferulic acid and sinapic acid, respectively. Pt4CL2 recombinant protein exhibited 100, 31, 26, 0 and 0% relative activity to 4-coumaric acid, caffeic acid, ferulic acid, 5-hydroxyferulic acid and sinapic acid, respectively. Neither recombinant protein showed detectable activity to sinapic acid.

The results of the tissue-specific expression pattern and substrate specificity suggests that in addition to the general function of 4CL, Pt4CL1 apparently is more related to lignin synthesis in the xylem tissue and Pt4CL2 apparently is more likely involved in flavonoid synthesis and UV protection.

It should be noted that the isolation and characterization of the Pt4CL1 and Pt4CL2 cDNA clones is described in Kawaoka et al., Proceedings of the 6th International Conference on Biotechnology in the Pulp and Paper Industry, Vienna, Austria (1995); and in Hu, Wen-Jing, Isolation and Characterization of 4-coumarate: Coenzyme A Ligase cDNAs and Genes from Quaking Aspen (*Populus tremuloides* Michx), Ph.D. Dissertation, Michigan Technological University, Houghton, Michigan (1997); and Tsai et al., Plant Physiol., 117, 101 (1998).

C. Transformation and Regeneration

Several methods for gene transformation of plant species with the 4CL sequence are available such as the use of *Agrobacterium*, electroporation, particle bombardment with a gene gun or microinjection.

Preferably, a 4CL cDNA clone is positioned in a binary expression vector in an antisense orientation under the control of double cauliflower mosaic virus 35S promoter. The vector is then preferably mobilized into a strain of *Agrobacterium* species such as *tumefaciens* strain C58/pMP90 and is used as the DNA delivery system due to its efficiency and low cost.

For example, with reference to Fig. 2, the binary expression pA4CL1 used for plant transformations is shown. Specifically, the Pt4CL1 cDNA is inserted in an antisense orientation into *Pac* I and *Bam*H I sites between the double CaMV 35S/AMV RNA4 and the 3′ terminator sequence of the nopaline synthase gene in a binary cloning vector pA4CL1 (Fig. 2). The binary vector containing hygromycin phosphotransferase (*HPT*) gene is modified from pBin 19. The gene construct pA4CL1 is available from Michigan Technological University, Institute of Wood Research, Houghton, Michigan.

The binary vector construct is mobilized into *Agrobacterium tumefaciens* using the freeze-thaw method of Holsters et al., Mol. Gen. Genet. 163: 181-187 (1978). For the freeze-thaw method, 1.5 ml of overnight cultures *Agrobacterium tumefaciens* strain C58/pMP90 is pelleted at 5000 x g for 3 minutes at 4°C and suspended in 1 ml of ice cold 20 mM $CaCl_2$. To the suspension is added 10 µl binary vector DNA (from an alkaline lysis minipreparation) and mixed by pipetting. The microcentrifuge tube is then frozen in liquid nitrogen for 5 minutes and thawed at 37°C for 5 minutes. After being cooled on ice, 1 ml of LB is added and the mixture is incubated at 28°C for 2 hours with gentle shaking. 200 µl of the cells is spread onto LB plates containing gentamycin and kanamycin and incubated at 28°C for 2 days. Colonies grown on the selection plates are randomly picked or miniprep and restriction enzyme digestion analysis is used to verify the integration.

The resulting binary vector containing *Agrobacterium* strain is used to transform quaking aspen according to Tsai et al., Plant Cell Rep. 14: 94-97 as set forth below.

Explants of young leaves from cuttings of aspen are obtained by cutting leaf disks of approximately 7 mm square from the young leaves along the midrib of the leaves. The explants are surface sterilized in 20% commercial bleach for 10 minutes followed by rinsing 3 times with sterile distilled, deionized water.

All of the culture media used includes the basal medium of woody plant medium (WPM) as described in Lloyd et al., Proc. Int. Plant Prop. Soc. 30: 421-437 (1980) and supplemented with 2% sucrose. 650 mg/L calcium gluconate and 500 mg/L MES are added as pH buffers as described in Tsai et al., Plant Cell Reports, 1994. All culture media is adjusted to pH 5.5 prior to the addition of 0.75% Difco Bacto Agar and then autoclaved at 121 °C and 15 psi for 20 minutes. Filter sterilized antibiotics are added to all culture media after autoclaving. All culture media are maintained at 23 ± 1 °C in a growth chamber with 16 hour photoperiods (160 $\mu E \times m^{-2} \times S^{-1}$) except for callus induction (as will be described later) which is maintained in the dark.

The sterilized explants are then inoculated with the mobilized vector with an overnight-grown agrobacterial suspension containing 20 μM acetosyringone. After cocultivation for 2 days, the explants are washed in 1 mg/ml claforan and ticarcillin for 2 hours with shaking to kill *Agrobacterium*. The explants are blotted dry with sterile Whatman No. 1 filter paper and transferred onto callus induction medium containing 50 mg/L kanamycin and 300 mg/L claforan to induce and select transformed callus. The callus induction medium is the basal medium with the addition of 6-benzyladenine (BA) and 2, 4-dichlorophenoxyacetic acid (2, 4-D) at concentrations of 0.5 mg/L and 1 mg/L, respectively, to induce callus.

The kanamycin-resistant explants are then subcultured on fresh callus induction media every two weeks. Callus formation occurs after approximately four weeks. Formed calli are separated from the explant and subcultured periodically for further proliferation.

When the callus clumps reach approximately 3 mm in diameter, the callus clumps are transferred to shoot regeneration medium. The shoot regeneration medium is the basal medium containing 50 mg/L kanamycin, 0.5 mg/L thidiazuron (TDZ) as a plant growth regulator and claforan at 300 mg/L to kill *Agrobacterium*. Shoots were regenerated about 4 weeks after callus is transferred to regeneration medium.

As soon as the shoots are regenerated, they are immediately transferred to hormone-free elongation medium containing 50 mg/L kanamycin and, whenever necessary, claforan (300 mg/L), to promote elongation. Green and healthy shoots elongated to 2-3 cm in length are excised and planted separately in a hormone-free rooting medium containing 50 mg/L kanamycin. The efficient uptake of kanamycin by shoots during their rooting stage provides the most effective selection for positive transformants. Transgenic plants are then transplanted into soil medium of vermiculite:peatmoss:perlite at 1:1:1 and grown in the greenhouse.

The above described transformation and regeneration protocol is readily adaptable to other plant species. Other published transformation and regeneration protocols for plant species include Danekar et al., Bio/Technology 5:587-590 (1987); McGranahan et al., Bio/Technology 6:800-804 (1988); McGranahan et al., Plant Cell Reports 8:512-616 (1990); Chen, Ph.D. Thesis, North Carolina State University, Raleigh, North Carolina (1991); Sullivan et al., Plant Cell Reports 12:303-306 (1993); Huang et al., In Vitro Cell Dev. Bio. 4:201-207 (1991); Wilde et al., Plant Physiol. 98:114-120 (1992); Minocha et al., 1986 Proc. TAPPI Research and Development Conference, TAPPI Press, Atlanta, pp. 89-91 (1986); Parsons et al., Bio/Technology 4:533-536 (1986); Fillatti et al., Mol. Gen. Genet 206:192-199 (1987); Pythoud et al., Bio/Technology 5:1323-1327 (1987); De Block, Plant Physiol. 93:1110-1116 (1990); Brasileiro et al., Plant Mol. Bio 17:441-452 (1991); Brasileiro et al., Transgenic Res. 1:133-141 (1992); Howe et al., Woody Plant Biotech., Plenum Press, New York, pp. 283-294 (1991); Klopfenstein et al., Can. J. For. Res. 21:1321-1328 (1991); Leple et al., Plant Cell Reports 11:137-141 (1992); and Nilsson et al., Transgenic Res. 1:209-220 (1992).

D. Phenotype Changes

The results of the transformation can be confirmed with conventional PCR and Southern analysis. Transferring 4CL cDNA in an antisense orientation down regulates 4CL in the plant. Expression of the 4CL has been found to be blocked up to 96 percent of 4CL enzyme activity in some transgenic plants.

In the aspen example, after acclimation, the transgenic aspen displayed an unusual phenotype, including big curly leaves, thick stem diameter, longer internodes, more young leaves in the shoot tip and a red pigmentation in the petioles extending into midvein leaves. Red coloration of the developing secondary xylem tissues is observed after peeling of the bark in the transgenic plants.

E. Accelerated Growth

Down regulation of 4CL altered growth of the transgenic plants. For example, transformation with an antisense 4CL sequence accelerated the growth of the plant. Enhanced growth is markedly noticeable at all ages. In particular the transgenic trees showed enhanced growth in the form of thicker stems and enlarged leaves as compared to control plants. These characteristics are retained in the vegetative propagules of these transgenic trees. Table 2 sets forth exemplary data with respect to several lines of transgenic quaking aspen grown in the greenhouse after eight months. Volume represents the overall quantitative growth of the plant.

THIS PAGE BLANK (USP

Table 2: Growth Measurement for Control and Transgenic Plants

Plant #	Height (cm)	Diameter (cm)*	Volume (cm^3)*	Average Length of Internode (cm)
Control 1	247.7	1.08	75.6	2.6
Control 2	250.2	1.01	66.8	2.8
11-1	304.8	1.15	105.5	3.1
11-2	248.9	1.01	66.4	3.4
11-3	241.3	0.84	44.6	3.2
11-4	288.3	0.94	66.7	3.1
11-5	246.4	0.92	54.6	3.3
11-7	226.7	1.13	75.7	3.4
11-8	289.6	1.16	102.0	3.3
11-9	287.0	1.76	232.6	4.3
11-10	252.7	0.83	45.6	3.1
11-11	247.7	0.86	48.0	3.5
12-1	247.7	1.1	78.4	2.7
12-2	199.4	0.96	48.1	2.5
12-6	294.6	0.92	65.2	3.2
16-1	227.3	0.95	53.7	2.7
16-2	278.1	0.97	68.5	3.4
16-3	265.4	1.09	82.5	3.5
17-2	243.8	0.89	50.5	2.6

* at 10 cm above ground

The averages for height, diameter, volume and average length between internodes for the control plants are as follows:

Height (cm)	248.95
Diameter (cm)	1.045
Volume (cm^3)	71.2
Ave. Length of Internodes (cm)	2.7

With respect to height alone, for those transgenic plants (11-1, 11-4, 11-8, 11-9, 12-6, 16-2, 16-3) having a statistically larger height than the control plants, the average height was 286.83 cm as compared to the control plant average height of 248.95 cm.

With respect to diameter alone, for those transgenic plants (11-1, 11-7, 11-8, 11-9) having a statistically larger diameter than the control plants, the average diameter was 1.30 cm as compared to the control plant average diameter of 1.045 cm.

With respect to volume alone, for those transgenic plants (11-1, 11-8, 11-9, 12-1, 16-3) having a statistically larger volume than the control plants, the average volume was 120.2 cm^3 as compared to the control plant average volume of 71.2 cm^3.

With respect to average length of internodes alone, for those transgenic plants (11-1, 11-2, 11-3, 11-4, 11-5, 11-7, 11-8, 11-9, 11-10, 12-6, 16-2, 16-3) having a statistically larger average length of internodes than the control plants, the average length of internodes was 3.39 cm as compared to the control plant average length of internodes of 2.7 cm.

As demonstrated in Table 2, while there are variations in growth among the transgenic plants, the average length of the internodes for the transgenic plants is consistently and significantly higher than that of the control plants. Moreover, there is also faster root initiation, and alterations, e.g., an increase, in root fresh weight and length, i.e., enhanced root growth. Variations in the growth of the transgenic plants is normal and to be expected. Preferably, a transgenic plant with a particular growth rate is selected and this plant is vegetatively propagated to produce an unlimited number of clones that all exhibit the identical growth rate.

F. Lignin

Down regulation of lignin pathway 4CL results in production of plants with reduced lignin content.

The following table shows the reduction of lignin content and 4CL enzyme activity in several transgenic aspen which were transformed with a homologous antisense 4CL sequence.

Table 3: Characterization of Transgenic Aspen Plants Harboring Antisense 4CL Sequence

Transgenic Plant #	Lignin Content % Based on Wood Weight	% Lignin Reduction	%4CL Enzyme Activity*	% 4CL Enzyme Activity Reduction
Control	21.4	0.0	868	0
11-1	20.5	4.2	1171	-25
11-2	19.2	10.3	515	45
11-3	20.4	2.3	922	6
11-4	19.7	7.9	1032	-19
11-5	19.7	7.9	691	20
11-7	19.9	7.0	578	38
11-8	20.2	5.6	694	20
11-9	20.4	4.7	806	14
11-10	19.4	9.3	455	51
11-11	20.4	4.7	726	22
12-1	12.8	40.2	49	95
12-2	12.6	41.1	62	93
12-3	11.9	44.4	61	94
12-6	19.8	7.5	786	16
16-1	12.8	40.2	35	96
16-2	20.6	3.7	780	17
16-3	21.0	1.9	795	15
17-1	20.5	4.2	855	9
17-2	21.4	0.0	925	1

*activity is expressed as pkat/(mg protein) using 4-coumaric acid as the substrate

Lignin content was determined according to Chiang and Funaoka (1990) Holzforschung 44:147-155. 4CL enzyme activity was determined according to Ranjeva et al. (1976), Biochimie 58:1255-1262.

The data in Table 3 demonstrates a correlation between down regulation of 4CL and reduction in lignin content. Transgenic plants with reduced lignin content have an altered phenotype in that the stem is more elastic to the touch or less curly.

It should also be noted that for those transgenic plants (12-1, 12-2, 12-3 and 16-1) with the approximately 40% reduction in lignin content and the corresponding approximately 95% reduction in 4CL enzyme levels, all of those transgenic plants had a consistent deep red coloration in the wood of the plant. Accordingly, the deep red color can be used as an identifier of reduced lignin content.

Down regulation of lignin pathway 4CL can also result in production of plants with an altered lignin structure. Based upon thioacidolysis (Rolando et al. (1992) Thioacidolysis, Methods in Lignin Chemistry, Springer-Verlag, Berlin, pp. 334-349) of plants 12-3 and 16-1, coniferyl alcohol and sinapyl alcohol lignin units are significantly reduced in these two plants as compared to the control tree, as shown in the following table.

Table 4: Altered Lignin Structure

Plant #	Coniferyl Alcohol Units *	Sinapyl Alcohol Units*
Control	733	1700
12-3	283	592
16-1	247	445

*micro-mole/g of lignin

The alteration of the frequency of the structural units in lignin of these transgenic plants is evidence that the overall structure of lignin in these plants has been genetically altered.

G. Cellulose Content

Down regulation of lignin pathway 4CL can result in increased cellulose content of the transgenic plants. Analysis of control and transgenic aspen for carbohydrate content demonstrate a higher cellulose content in the transgenic plants than the control plants. Particularly, the transgenic plants that have over 40% lignin reduction have about 10-15% higher cellulose content than the control. Data is set forth in the following tables for trees that were transformed with homologous 4CL in an antisense orientation:

Table 5: Analysis of Carbohydrate Components in Transgenic and Control Aspen

Plant #	Glucan	Arabinan	Galactan	Rhamnan	Xylan	Mannan
Control	44.23%	0.47%	0.79%	0.37%	17.19%	1.91%
11-2	49.05%	0.36%	1.05%	0.38%	15.34%	2.04%
11-9	45.95%	0.40%	0.80%	0.37%	17.12%	1.83%
11-10	47.49%	0.43%	0.99%	0.40%	16.24%	2.35%
12-3	50.83%	0.55%	1.24%	0.48%	17.25%	1.77%
16-1	48.14%	0.56%	1.07%	0.48%	19.14%	1.58%
16-2	46.55%	0.34%	0.82%	0.37%	16.75%	2.31%

Table 6: Comparison of Lignin and Cellulose (glucan) Contents in Transgenic and Control Aspen

Plant #	Lignin		Cellulose	
	Content % on Wood	% Reduction	Content % on Wood	% Increase
Control	21.4	0	44.23	0
11-2	19.2	10.3	49.05	10.9
11-9	20.4	4.7	45.95	3.9
11-10	19.4	9.3	47.49	7.4
12-3	11.9	44.5	50.83	14.9
16-1	12.8	40.2	48.14	8.8
16-2	20.6	3.7	46.55	5.2
11-6	18.6	13.1	45.98	3.8
12-1	12.5	40.2	48.35	9.3
12-2	12.6	41.1	49.74	12.5
12-5	14.4	32.7	45.58	3.1

The procedure for carbohydrate analysis utilized is as follows. About 100 mg of milled woody tissue powder with sizes that pass a 80-mesh screen was hydrolyzed with 1 mL of 72% (W/W) H2SO4 for 1 hr at 30°C. Samples were then diluted to 4% (W/W) H2SO4 with distilled water, fucose was added as

THIS PAGE BLANK (USP

an internal standard, and a secondary hydrolysis was performed for 1 hr at 121°C. After secondary hydrolysis, the sugar contents of the hydrolysates are determined by anion exchange high performance liquid chromatography using pulsed amperometric detection. Sugar contents are expressed as % of the weight of the woody tissue used. The above procedures are similar to those in a publication by Pettersen and Schwandt, J. Wood Chem & Technol. 11:495-501 (1991).

H. Increased Disease Resistance

Down regulation of lignin pathway 4CL can result in altered levels of phenylpropanoids or secondary metabolities that display antimicrobial activity. Thus, transgenic plants with down-regulated 4CL can result in enhanced disease resistance, and in particular, with increased fungal pathogen resistance. In particular, greenhouse transgenic aspen plants may show a disease resistance to fungi such as those which induce leaf-blight disease.

I. Promoters

Two distinct genes encoding 4CL and their promoters were cloned. The promoter of Pt4CL1 can drive gene expression specifically in xylem tissue and the promoter for Pt4CL2 confers gene expression exclusively in the epidermal tissues. These promoters can be used to manipulate gene expression to engineer traits of interest in specific tissues of target plants. The significance of the promoters is the application of the xylem-specific promoter to direct the expression of any relevant genes specifically in the xylem for engineering lignin content, lignin structure, enhanced growth, cellulose content, other value-added wood qualities, and the like. The importance of the epidermis-specific promoter is its ability to drive the expression of any relevant genes specifically in epidermal tissues for engineering disease-, UV light-, cold-, heat-, drought-, and other stress resistance traits in plants.

Specifically, the promoters of the Pt4CL1 and Pt4CL2 were isolated as follows. An aspen genomic library was screened with Pt4CL1 cDNA and Pt4CL2 partial cDNA fragment to isolate genomic clones of Pt4CL1 and Pt4CL2. Eleven and seven positive genomic clones were identified for Pt4CL1 and Pt4CL2 gene, respectively. Among 11 positive clones for Pt4CL1,

THIS PAGE BLANK (USPTO)

Pt4CL1g-4 contained a full length coding sequence and at least 2 kb of 5′ flanking regions. The restriction map of Pt4CL1g-4 is set forth at Fig. 3.

With respect to Pt4CL2, restriction map analysis was performed on λDNA of positive genomic clone Pt4CL2g-11 which contains a full length coding sequence and about 1.2 kb of 5' flanking region. The restriction map of Pt4CL2g-11 is set forth at Fig. 4.

Approximately a 2.3 kb 5′ flanking region of Pt4CL1 was digested from Pt4CL1g-4 using *Xba* I and *Sac* I sites and cloned into pGEM7Z *Xba* I and *Sac* I sites. The subcloned Pt4CL1 promoter was named p7Z-4XS and the restriction map of P7Z-4XS is set forth at Fig 5. The 5′ unilateral deletion of p7Z-4XS was generated for DNA sequencing by exonuclease III/S1 nuclease digestion using Erase-a-Base System (Promega, Madison, WI). The deletion series was sequenced using a primer on pGEM7Z vector.

A 1.5 kb *Hind* III and *Eco*R I fragment containing a 1.2 kb 5′ flanking region of Pt4CL2 and 0.3 kb coding region of Pt4CL2g-11 was subcloned in pBluescript II SK+ *Hind* III and *Eco*R I sites. The restriction map of the resulting clone, pSK-11HE, was determined by digesting the plasmid with several restriction enzymes, as in set forth at Fig. 6. In order to determine the sequence of the Pt4CL2 promoter, pSK-11HE was further digested into small fragments according to the restriction map and subcloned into vectors with suitable cloning sites. The DNA sequence was determined using M13 universal primer and reverse primer on the vector.

The DNA sequences of the two promoters was determined and analyzed using ΔTaq cycle sequencing Kit (USB, Cleveland, OH), and GENETYX-MAC 7.3 sequence analysis software from Software Development Co., Ltd. The nucleotide sequence of promoter region of Pt4CL1 is set forth as SEQ ID NO:5 and the nucleotide sequence of the promoter region of Pt4CL2 is set forth as SEQ ID NO:6. The sequence of the promoter regions of Pt4CL1p and Pt4CL2p is available from Genbank, Accession Nos. AF041051 and AF041052, respectively.

The insignificant sequence similarity between the 5′- and 3′-noncoding regions of these two genes and their distinct exon-intron organizations (four

introns in Pt4CL1 and five in Pt4CL2) further substantiate their functional and perhaps evolutionary divergency. Striking differences also were observed in the promoter sequences of these two genes. Three cis-acting elements, box P (CCTTTCACCAACCCCC; SEQ ID NO:15), box A (CCGTTC; SEQ ID NO:16), and box L (TCTCACCAACC; SEQ ID NO:17), previously shown to be consensus in all known plant phenylalanine ammonialyase (PAL) and 4CL gene promoters (Hahlbrook et al., Proc. Natl. Acad. Sci. USA, 92, 4150 (1995); Logemann et al., Proc. Natl. Acad. Sci. USA, 92, 5905 (1995)), were identified within the 1kb 5′ flanking sequence of Pt4CL1 (GenBank Accession No. AF041051). However, none of these boxes could be found within the analyzed 1.2 kb 5′ flanking region of Pt4CL2 (GenBank Accession No. AF041052), suggesting that promoter differences between Pt4CL1 and Pt4CL2 genes could be responsible for the strikingly different patterns of tissue-specific expression of these genes, as observed in Northern analysis.

Tissue-specific expression can be achieved by fusing the promoters of Pt4CL1 or Pt4CL2 to a gene, e.g., an open reading frame of interest and transferred to a plant species via Agrobacterium. For the sake of example, the promoters of Pt4CL1 and Pt4CL2 were fused to a GUS reporter gene as detailed below. However, it should be noted that genes other than the GUS reporter gene can be fused to these promoters for tissue specific expression.

In order to construct Pt4CL1 promoter-GUS binary vector, a 1 kb fragment covering 5′-flanking region and 17 bp coding region of Pt4CL1 was subcloned into pGEM7Z *Sph* I and *Eco*R I sites for constructing promoter-GUS binary vector. In this 1 kb DNA fragment, it is found that one *Xho* I site is located at 486 bases upstream to the translation start site and the *Eco*R I site is located at 17 bases downstream the translation start site. This 0.6 kb fragment was subcloned into pGEM7Z *Xho* I and *Eco*R I sites and used as a template in PCR amplification.

In order to construct a promoter-GUS transcriptional fusion, a *Bam*H I site was introduced in front of the translation start site of Pt4CL1 by PCR. PCR amplification was performed using p7Z-4XE as the template, M13 universal primer on pGEM7Z vector as 5′ end primer and Pt4CL1p-1 primer containing a

*Bam*H I site at the end is complementary to a sequence upstream of the translation start site. The reaction was carried out in 100 µl reaction mix containing 1x *pfu* reaction buffer, 200 µl each dNTPs, 100 µM each primer and 5 units of *pfu*. The PCR reaction mixture was denatured at 94°C for 5 minutes followed by 30 cycles of 94°C (1 minute), 55°C (1 minute), 72°C (1 minute, 30 seconds) and was ended with a 5 minute extension at 72°C.

The amplified 0.6 kb fragment was cloned and sequenced to confirm the sequence. The engineered 0.6 kb fragment was ligated to p7Z-4SE which was digested with *Xho* I and *Bam*H I. In order to incorporate a *Hind* III site in the 5′ end of Pt4CL1 promoter, the 1 kb *Sph* I-*Bam*H I Pt*C*CL1 promoter region was the cloned into pNoTA (5 prime → 3 prime Inc., Boulder, CO) *Sph* I and *Bam*H I site. The 1 kb Pt4CL1 promoter was then released from pNoTA vector with *Hind* III and *Bam*H digestion and subsequently transcriptionally fused to pBI101 *Hind* III and *Bam*H I sites in front of GUS. The resulting binary vector was named Pt4CL1p-GUS and is set forth at Fig. 7.

In order to construct Pt4CL2 promoter-GUS binary vector, pSK-11HE was digested with *Sph* I and *Eco*R I to release 0.2 kb *Sph* I and *Eco*R I fragment. The 0.2 kb fragment was cloned into pGEM7Z *Sph* I and *Eco*R I sites. A primer, Pt4CL2p-3′ (5′-CATCGGATCCTGAGATGGAAGGGAGTTTCT-3′; SEQ ID NO:15) was designed to be complementary to a sequence upstream of the translation start site of Pt4CL2 and to incorporate *Bam*H I site at the end. Amplification was performed using p7Z11SE as a template, M13 universal primer as the 5′ end primer and Pt4CL2p-3 as the 3′ end primer. A PCR reaction was carried out and the amplified PCR product was cloned and sequenced to check the fidelity of the PCR amplification. The 0.2 kb *Sph* I-*Bam*H I DNA fragment with correct sequence was fused to pSK-11HE linearized with *Sph* I and *Bam*H I. The resulting plasmid was named pSK-11HB. The promoter of Pt*C*CL2 was then excised from pSK-11HB with *Hind* III and *Bam*H I and ligated to PBI101 and *Hind* III and *Bam*H I site to make Pt4CL2p-GUS transcriptional fusion binary vector as shown in Fig. 8.

The Pt4CL1p-GUS and Pt4CL2p-GUS constructs were then mobilized into *Agrobacterium tumefaciens* strain C58/pMP90 by freeze and thaw method as explained previously.

Leaf disk transformation of tobacco with these two Agrobacterium constructs is conducted according to the method of Horsch R.B. (1988) Leaf Disk Transformation, Plant Molecular Biology Manual, A5:1-9.

To further investigate the regulation of the tissue-specific expression of Pt4CL1 and Pt4CL2 genes at the cellular level, their promoter activities were analyzed in transgenic tobacco plants by histochemical staining of GUS gene expression driven by a 1 kb Pt4CL1 and 1.2 kb Pt4CL2 promoter sequences, respectively. In Pt4CL1p-GUS transgenic plants, intense GUS staining was detected in lignifying xylem of stem. Strong GUS activity also was found localized to xylem of leaf mid-rib and of root. However, there was no GUS expression in leaf blade, stem epidermis, cortex, phloem and pith, and flower petal. These results are consistent with the evidence that Pt4CL1 gene expression is xylem- or lignifying tissue-specific, and with the observation that Pt4CL1 mRNA level is the highest in aspen secondary developing xylem. In striking contrast to the Pt4CL1 promoter activity, the Pt4CL2 promoter did not direct GUS expression in vascular and xylem tissues in the stem and the leaf of Pt4CL23p-GUS transgenic plants. Instead, it directed GUS expression in lignin-deficient epidermal cells of the stem (Figure 10C) and of the leaf, reflecting the association of Pt4CL2 with nonlignin-related phenylpropanoid biosynthesis in the plant's outer layers. In addition, GUS staining also was detected in Pt4CL2p-GUS transgenic plant's floral organs, such as stigma and petal, suggesting the likely relevance of Pt4CL2 in mediating the formation of flavonoids, which are known to be accumulated in these organs (Higuchi (1997, supra; Caldwell et al., Physiol. Plant, 58, 455 (1983); Shirley, Trends in Plant Sci., 1, 377 (1996)).

The epidermis-specific Pt4CL2 promoter activity indicated that the in vivo Pt4CL2 mRNA message observed in aspen stem internodes could be caused by the signal derived from the epidermis RNA. Thus, the specific expression of

Pt4CL2 mRNA in epidermis further supports the biochemical functions of Pt4CL2 protein in the biosynthesis of nonlignin-related phenylpropanoids.

Therefore, the promoter fragments incorporated in Pt4CL1p-GUS and Pt4CL2p-GUS fusion genes must encompass the regulatory sequence elements that are responsible for the contrasting tissue-specific expression between Pt4CL1 and Pt4CL2 genes in aspen. Thus, based on both in vivo gene expression and gene promoter activity analyses, it was concluded that the expression of Pt4CL1 and Pt4CL2 genes in aspen is compartmentalized.

These results demonstrate that in aspen two functionally distinct 4CLs are uniquely compartmentalized by their gene regulatory systems for mediating differentially the biosynthesis of lignin and other phenylpropanoids that serve different physiological functions in aspen. Pt4CL1 is involved in channeling hydroxycinnamic acid derivatives to the synthesis of guaiacyl-syringyl lignin in xylem tissues. Pt4CL2 is associated with the biosynthesis of phenylpropanoids other than lignin in epidermal cells in the stem and the leaf, suggesting its likely participation in disease-resistance or defense-related mechanisms in the plant's outer layers. Therefore, 4CL isoforms may have distinct roles in plant defense systems and in lignification in a tissue-specific manner. From a practical point of view, the tissue-specific Pt4CL1 and Pt4CL2 gene promoters may offer a more defined control of future genetic engineering of traits in trees that must be confined to xylem or epidermal cells.

J. Cellulose Accumulation

Twenty-five transgenic aspen lines were generated in which Pt4CL1 expression was down-regulated to various degrees by antisense inhibition, using a Pt4CL1 gene operatively linked to a duplicated enhancer CaMV 35S promoter (Datla et al., Plant Sci., 94, 139 (1993)). The effect of Pt4CL1 deficiency on woody tissue development was investigated in ten-month-old trees. Pt4CL1 messenger RNA was drastically reduced in four lines (Fig. 9A). These lines also exhibited more than a 90% reduction in xylem Pt4CL1 enzyme activity (Fig. 9B), and a 40 to 45% reduction in stem lignin (Fig. 9C). A more modest lignin reduction was found in those lines with less drastic repression of Pt4CL1 activity. The reduction in lignin content was restricted to woody xylem, as

shown by attenuated lignin autofluorescence in xylem but not in phloem fibers following UV-irradiation (Figs. 9D, E). Severe repression of other lignin biosynthetic pathway enzymes, such as COMT or CAD, had no effect on lignin quantity in transgenic aspen, hybrid poplar or a loblolly pine (Pinus taeda) mutant (Tsai et al., 1998; VanDoorsselaere et al., Plant J., 8, 855 (1995); Baucher et al., Plant Physiol., 112, 1479 (1996)). Lignin structure, however, was significantly altered in these cases.

To investigate the effect of Pt4CL1 repression on lignin structure, milled wood lignins were isolated from the stem of a transgenic (line 6 with a 45% lignin reduction) and a control (using methods described in Bjorkman, Nature, 174, 1057 (1954); Chiang et al., Holzforschung, 44, 147 (1990); and Ralph et al., JACS, 116, 9448 (1994)) and then were analyzed by nuclear magnetic resonance (NMR) Examination of HSQC (heteronuclear single-quantum coherence) spectra (Fig. 10) and their HSQC-TOCSY (HSQC-total correlation spectroscopy) counterparts and HMQC (heteronuclear multiple-quantum correlation) indicated that these lignins are structurally similar, consistent with their comparable syringyl-to-guaiacyl ratios based on thioacidolysis of intact stem. The ratios for control and transgenic line 6 were 2.3 and 2.1, respectively. Thus, there appeared to be little disruption of the normal lignin structure as a result of reduced Pt4CL1 activity. It is clear from Figure 10 that β-aryl ethers (β-O-4) **10**, normally the most abundant (50 to 60%) linkage type in tree lignin (Adler et al., Wood Sci. Technol., 11, 169 (1977)), predominate in both lignin samples. In both lignins, erythro-isomers are more prevalent than their threo-counterparts, typical of angiosperm lignin. Resinol (β-β) units (**12** Fig. 10), which largely results from coupling of sinapyl alcohol **9b** monomers and represent initial intermediates in lignin polymerization reactions in angiosperm trees, are well represented in both lignins. Traces of phenylcoumaran (β-5) units **11** and α-β-diaryl ethers **14** were detectable in each lignin. Absent from both lignins were condensed biphenyl units such as dibenzodioxocins **13** (Ralph et al., supra). Such units, formed from 5-5-homo-coupling of coniferyl alcohol **9a**, normally represent about 4% of the constituents in angiosperm lignin (Adler, supra).

Low levels of 4-coumaric **2** and ferulic **4** acids are sometimes detectable in angiosperm lignins. Therefore, it was determined whether the incorporation of these acids was affected by decreased Pt4CL1 activity. Long-range ^{13}C-^{1}H-correlation (HMQC) NMR experiments revealed that these acids were absent from both lignin samples. However, cell walls of transgenic stem tissue contained alkaline extractable 4-coumaric **2** and ferulic **4** acids at levels 11- and 5-fold higher, respectively, than the control. Alkaline hydrolysis of stem wood meal (pass 80-mesh) was performed at room temperature for 24 hr in 1 N NaOH (Hartley, J. Chromatogr., 54, 335 (1971)). The hydrolysates were neutralized, extracted with ethyl acetate and concentrated. The concentrated products were derivatized with BSTFA and analyzed by GC-MS in SIM (selected ion monitoring) mode using a DB-5 column. 4-Coumaric acid **2** (TMS-derivative; m/z 308) content of control was 199 ± 13 nmol/g dry wood, and 2145 ± 93 nmol/g dry wood in transgenic line 6. Ferulic acid **4** (TMS-derivative: m/z 338) contents in control and transgenic line 6 were 510 ± 9 and 2431 ± 120 nmol/g dry wood, respectively. No sinapic acid 6 (TMS-derivative: m/z 368) could be detected in control. However, a significant amount of sinapic acid, 2452 ± 119 nmol/g dry wood, was found in transgenic line 6.

Together, the lignin and cell wall analyses support a requirement for activation by Pt4CL1 of these phenolic acids for their incorporation into lignin. The cell wall apparently serves as a sink for accumulating these acids when Pt4CL1 activity is reduced. As a result, lignin content was reduced in the transgenic line but lignin composition and structure were not significantly altered. The conservation of normal lignin composition and structure in the transgenic aspen stands in sharp contrast to the marked changes of lignin composition and structure in other transgenic and mutant plants with altered lignin biosynthesis (Tsai et al., 1998; Van Doorsselaere et al., 1995; Baucher et al., 1996; Elkind et al., Proc. Natl. Acad. Sci. USA, 87, 9057 (1990); Piquemal et al., Plant J., 13, 17 (1998); Sewalt et al., Plant Physiol., 115, 41 (1997); Kajita et al., Plant Physiol., 114, 871 (1997); Lee et al., Plant Cell, 9, 1985 (1997); Dwivedi et al., Plant Mol. Biol., 26, 61 (1994); Ni et al., Transgenic Res., 3, 120 (1994); Atanassova et al., Plant J., 8, 465 (1995); Halpin et al., Plant J., 6, 339

(1994); Hibino et al., Biosci. Biotech. Biochem., 59, 929 (1995)). The results are consistent with the supposition that 4CL modulates lignin accumulation in trees in a regulatory manner that does not result in disruption of lignin structure.

Lignin and polysaccharides are proposed to account for the remarkable mechanical strength of woody tissues (White et al., Nature, 205, 818 (1965); Atalla et al., Science, 227, 636 (1985); Houtman et al., Plant Physiol., 107, 977 (1995); Taylor et al., Plant J., 2, 959 (1992); Turner et al., Plant Cell, 9, 689 (1997)). In consideration of the possible effects of severe lignin reduction on structural polysaccharide components, these components were examined in stem wood tissue. While hemicellulose content remained essentially unchanged, the transgenic lines had a 9 to 15% increase in glucan (Table 7), identified as β-(1→4)-glucan, or cellulose, by methylation-based linkage analysis and enzymatic hydrolysis. Lignin content was determined as the sum of Klason and acid-soluble lignins which represent the absolute quantity of lignin (Chiang et al., Holzforschung, 44, 147 (1990)). Cellulose and hemicelluloses contents were determined based on the total sugars after acid hydrolysis of these polysaccharides in stem woody tissue (Chiang et al., Wood Sci. Technol., 17 , 217 (1983); Pettersen et al., J. Wood Chem. Technol., 11, 495 (1991)). Wood meal (pass 80-mesh) was vacuum-dried at 45°C and hydrolyzed with H_2SO_4. Sugar contents of the hydrolysates were determined by anion exchange high performance liquid chromatography using pulsed amperometric detection and used for quantifying glucan and other polysaccharides (hemicelluloses) (Davis, J. Wood Chem. Technol., 18, 235 (1998)).

The dried wood meal was also used for methylation analysis of the glucan in wood. Both the Hakomori (J. Biochem. Tokyo, 55, 205 (1964)) and $NaOH/CH_3I$ (Ciucanu et al., Carbohydr. Res., 131, 209 (1984)) methylation procedures were followed. Methylated samples were hydrolyzed in 2M TFA at 121°C for 2 hr, reduced with sodium borodeuteride, and acetylated using acetic anhydride at 120°C for 3 hr. The derivatized samples were analyzed by GC-MS using a Sp2330 Supelco column. The methylation revealed that the glucose residues are mainly derived from 1→4 glucan for both control and transgenic

lines. Enzymatic hydrolysis of stem woody tissue further confirmed that the glucans in both control and transgenic lines are β-(1→4)-glucan, or cellulose.

Thus, (1→3)-linked glucan (callose), reportedly deposited in plant cell walls as a result of perturbed secondary metabolism (Schmelzer et al., Plant Cell, 1, 993 (1989)), was not detected in transgenic or control wood. Together, increased cellulose and decreased lignin content resulted in a cellulose-to-lignin ratio of 4 compared with 2 in control aspen (Table 7). The reason for the increased cellulose content is not clear. The absence of change in transcript levels of an aspen homolog of celA encoding a catalytic subunit of cellulose synthase (Arioli et al., Science, 279, 717 (1997)) argues against an increase in the rate of cellulose deposition due to altered transcriptional regulation in transgenic trees with reduced lignin content. The increase in cellulose content suggests that cross talk between lignin and cellulose biosynthetic pathways can nevertheless occur to ensure that cellulose biosynthesis becomes the preferred structural carbon sink when lignin biosynthesis is reduced. Because cellulose and lignin are the two components of wood most responsible for its rigidity, such cross talk could represent an adaptation to sustain mechanical strength in lignin deficient xylem.

The reduced lignin content in transgenic lines did not adversely affect tree growth and development. In fact, trees with down-regulated Pt4CL1 had thicker stems, longer internodes, and larger (frequently epinastic) leaves than controls (Figs. 11A and 11B). Scanning electron microscopy (SEM) revealed that the shape and size of stem xylem fiber and vessel cells were similar to those of controls (Figs. 11C-F). Therefore, the enhanced stem development in these transgenic lines was apparently due to increased proliferative activity during xylem development rather than to increased cell size. Root growth rates also increased in these lines, resulting in greater length (15-fold) and fresh weight gain (20-fold) than in controls over a 14-day period in ex vitro rooting experiments (Fig. 11G). Cell size distribution in the meristematic and elongation zones of root tips was similar in control and transgenic roots. As was the case in stem xylem, increased root growth rate of the transgenic was due to increased cell number. Leaf growth also increased in the transgenic lines resulting in 4- to

5-fold larger leaves than in controls (Fig. 11B). Mature leaf adaxial epidermal cells were measured in two of the transgenic lines and found to be at least twice as large as in control aspen. A more detailed analysis was conducted to determine whether the rate and/or the duration of cell expansion accounted for the increased cell size in mature leaves of transgenic aspen. Epidermal cell expansion stopped at leaf number 15 below the first emerging leaf in control plants, but epidermal cells as well as leaf area continued to expand at leaf number 28 in transgenics (Fig. 11H). Therefore, the prolonged expansion of epidermal cells contributed to increased leaf size in the transgenic aspen lines.

The promotive effects on growth and development in the transgenic trees was a surprising observation. Growth enhancement has not been reported in transgenic tobacco or Arabidopsis with downregulated PAL (phenylalanine ammonia lyase), CCR, C4H, 4CL, COMT, or CAD. In fact, stunted growth and collapsed cell walls occurred in some transgenic tobacco with altered lignin biosynthesis. Whether the growth responses between herbaceous and tree species differed due to altered lignin biosynthesis per se is not clear. In the case of aspen, lignin composition and structure were conserved, eliminating the possibility that altered lignin constituents promoted growth. In aspen trees, reduced expression of Pt4CL1 disrupted lignin biosynthesis downstream of the phenylpropanoid pathway and this increased the concentration of phenylpropanoid intermediates in cell walls. At the same time, enhanced cell division and cell expansion were observed in root tips and leaves. Whether the growth enhancement observed in the transgenic aspen is due to altered carbon distribution between primary/secondary metabolism or specifically due to changes in wall-bound moieties are two possibilities to consider. Histone gene(s) expression has been used as a marker to show that cell division decreases in suspension cells and young leaves of parsley following treatments of that divert carbon flow in to the phenylpropanoid pathway and away from primary metabolic pathways (Logemann et al., Plant J., 8, 865 (1995)). There is also current interest in the organization and composition of cell wall constituents and their effects on cell expansion and plant growth. For these rationale,

phenylpropanoid flux as well as cell wall constituents would be of interest for investigating growth effects of lignin manipulation in trees.

The finding that cellulose content increases in transgenic aspen with disrupted lignin biosynthesis is unique; similar observations have not been reported in herbaceous plants (Turner et al., Plant Cell, 9, 689 (1997); Elkind et al., 1990; Piquemal et al., 1998)). Interesting to consider is the idea that in perennial woody plants, lignin and cellulose deposition in cell walls are regulated in a compensatory fashion such that decreased in one are compensated for by increases in the other for maintaining the cellular structural integrity. This compensatory deposition of lignin and cellulose is consistent with the manner of how trees regulate their lignin and cellulose quantities in the course of forming naturally occurring reaction wood for mechanical support. Compensatory regulation such as this would also provide metabolic flexibility during annual growth increments, perhaps key for the long term structural integrity of woody perennials like trees. Further study is required to determine whether such regulation of cellulose accumulation is sensitive to primary/secondary metabolism and to changes in cell wall constituents such as those observed in Pt4CL1 down-regulated aspen.

Overall, lignin limits the utilization of wood for fiber/material-, chemical-, and energy-production. Traditional breeding approaches have not led to trees with more desirable lignin/cellulose composition. However, genetic engineering appears to offer a strategy for manipulating such traits in trees, with the prospect of systemically regulating growth as reported here. The benefit of these engineered traits may also extend to forage crops in which lignin has been identified as the major barrier to their digestibility by ruminants.

Table 7. Lignin and cellulose contents in stem woody tissue of control and transgenic aspen. Data are the mean ± SD of three independent experiments. Normalized values relative to control are shown in parentheses.

	Lignin Content	Cellulose Content	
Line	(% of dry wood weight)	(% of dry wood weight)	Cellulose-to-lignin ratio
Control	21.62 ± 0.30 (100)	44.23 ± 0.43 (100)	2.0
4	12.83 ± 0.28 (60)	48.35 ± 0.60 (109)	3.8
5	13.02 ± 0.28 (60)	49.74 ± 0.45 (112)	3.7
6	11.84 ± 0.08 (55)	50.83 ± 0.26 (115)	4.3
8	12.90 ± 0.04 (60)	48.14 ± 0.29 (109)	3.8

All publications and patents are incorporated by reference herein, as though individually incorporated by reference, as long as they are not inconsistent with the present disclosure. The invention is not limited to the exact details shown and described, for it should be understood that many variations and modifications may be made while remaining within the scope of the invention defined by the claims.

WHAT IS CLAIMED IS:

1. A method for altering the growth characteristics of a plant comprising the step of incorporating into the genome of the plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase such that when the nucleotide sequence is expressed in the plant, the growth of the plant is altered.

2. The method as set forth in claim 1 wherein the DNA molecule comprises a homologous nucleotide sequence.

3. The method as set forth in claim 1 wherein the DNA molecule comprises a heterologous nucleotide sequence.

4. The method as set forth in claim 1 wherein the DNA molecule is incorporated into the genome of the plant by transformation using an Agrobacterium transfer vector.

5. The method asset forth in claim 1 wherein the DNA molecule comprises the nucleotide sequence in antisense orientation.

6. The method asset forth in claim 1 wherein the DNA molecule comprises the nucleotide sequence in sense orientation.

7. The method as set forth in claim 6 wherein the DNA molecule is a cloned cDNA sequence of 4-coumarate Co-enzyme A ligase.

8. The method as set forth in claim 1 wherein the recombinant DNA molecule comprises the promoter sequence of CaMV35S.

9. The method as set forth in claim 1 wherein said altered growth is manifested as an increase in biomass.

10. A plant having its growth characteristic altered through the incorporation into the genome of the plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase such that when the nucleotide sequence is expressed in the plant, the growth characteristic of the plant is altered.

11. The plant as set forth in claim 10 wherein the DNA molecule comprises the nucleotide sequence in antisense orientation.

12. The plant as set forth in claim 10 wherein the DNA molecule comprises the nucleotide sequence in sense orientation.

13. The plant as set forth in claim 10 wherein the DNA is incorporated into the genome of the plant by transformation using an Agrobacterium transfer vector.

14. The plant as set forth in claim 10 wherein the DNA molecule is a cloned cDNA sequence encoding 4-coumarate Co-enzyme A ligase.

15. The plant as set forth in claim 10 wherein the DNA molecule comprises the promoter of CaMV35S.

16. The plant as set forth in claim 10 wherein said altered growth is manifested as an increase in plant biomass.

17. The plant as set forth in claim 10 which is a tree.

18. A method for altering a characteristic of a plant comprising the step of genetically down regulating the enzyme 4-coumarate Co-enzyme A ligase, wherein the characteristic is selected from the group consisting of accelerated growth, reduced lignin content, altered lignin structure, increased disease resistance and increased cellulose content.

THIS PAGE BLANK (USP

19. The method of claim 18 wherein the plant is genetically down regulated through incorporation into the genome of the plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase in antisense orientation.

20. The method as set forth in claim 18 wherein the recombinant DNA molecule is incorporated into the genome of the plant by transformation using an Agrobacterium transfer vector.

21. The method as set forth in claim 19 wherein the recombinant DNA molecule comprises a homologous nucleotide sequence that is incorporated into the genome of the plant.

22. The method as set forth in claim 18 wherein the nucleotide sequence is a cloned cDNA sequence encoding 4-coumarate Co-enzyme A ligase.

23. The method as set forth in claim 18 wherein the recombinant DNA molecule comprises a promoter of CaMV35S.

24. A plant having a characteristic altered by genetically down regulating the enzyme 4-coumarate Co-enzyme A ligase, wherein the characteristic is selected from the group consisting of accelerated growth, reduced lignin content, altered lignin structure, increased disease resistance and increased cellulose content.

25. The plant of claim 24 wherein the plant is genetically down regulated through incorporation into the genome of the plant a recombinant DNA molecule comprising a homologous nucleotide sequence encoding 4-coumarate Co-enzyme A ligase in the antisense orientation.

26. The plant of claim 24 wherein the recombinant DNA molecule is incorporated into the genome of the plant by transformation using an Agrobacterium transfer vector.

27. The plant of claim 24 wherein the nucleotide sequence is a cloned cDNA sequence of 4-coumarate Co-enzyme A ligase.

28. The plant of claim 24 wherein the recombinant DNA molecule comprises a promoter of CaMV35S.

29. An isolated and purified DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase gene.

30. The isolated and purified DNA molecule of claim 29 in which the DNA segment is from aspen.

31. The isolated and purified DNA molecule of claim 29 wherein the DNA segment directs expression of a linked sequence to the xylem of a plant.

32. The isolated and purified DNA molecule of claim 29 wherein the DNA segment directs expression of a linked sequence to the epidermal tissue of a plant.

33. A method of imparting disease resistance to a plant tissue comprising:

(a) introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase operably linked to a promoter functional in a plant cell into cells of a plant;

(b) regenerating said plant cells to provide a transgenic plant; and

(c) expressing the recombinant DNA molecule in the cells of the transgenic plant in an amount effective to render the plant resistant to disease.

34. The method according to claim 33 wherein the disease is caused by a fungus.

35. The method according to claim 33 wherein the nucleotide sequence is in the antisense orientation.

36. A transgenic plant prepared by the method of claim 33.

37. A transgenic seed of the transgenic plant of claim 33.

38. A transgenic plant, which plant is substantially resistant to disease, comprising:

(a) a native 4-coumarate Co-enzyme A ligase gene, and

(b) a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase operably linked to a promoter functional in a plant wherein the recombinant DNA molecule is expressed in an amount effective to confer resistance to the transgenic plant.

39. A method for altering the lignin content in a plant comprising:

(a) introducing an expression cassette comprising a recombinant DNA molecule encoding a 4-coumarate Co-enzyme A ligase operably linked to a promoter functional in a plant cell into the cells of a plant;

(b) regenerating said plant cells to provide a transgenic plant; and

(c) expressing the recombinant DNA molecule in the cells of the transgenic plant in an amount effective to alter the lignin content in the plant cells.

40. A transgenic plant having an altered lignin content in the plant cells comprising: a recombinant DNA molecule comprising a nucleotide sequence encoding a plant 4-coumarate Co-enzyme A ligase operably linked to a promoter so that the recombinant DNA molecule is expressed in an amount effective to alter the lignin content of the plant.

41. A method for altering the cellulose content in a plant comprising:

(a) introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase operably linked to a promoter functional in a plant cell into the cells of a plant;

(b) regenerating said plant cells to provide a transgenic plant; and

(c) expressing the recombinant DNA molecule in the cells of the transgenic plant in an amount effective to alter the cellulose content in the plant.

42. A transgenic plant having an altered cellulose content in the plant cells comprising: a recombinant DNA molecule comprising a recombinant DNA molecule sequence encoding a plant 4-coumarate Co-enzyme A ligase operably linked to a promoter so that the recombinant DNA molecule is expressed in an amount effective to alter the cellulose content of the plants.

43. A method for altering the lignin structure in a plant comprising:

(a) introducing an expression cassette comprising a recombinant DNA molecule comprising a recombinant DNA nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase operably linked to a promoter functional in a plant cell into the cells of a plant;

(b) regenerating said plant cells to provide a transgenic plant; and

(c) expressing the recombinant DNA molecule in the cells of the transgenic plant in an amount effective to alter the lignin structure in the plants.

44. A transformed plant having an altered lignin structure comprising: a recombinant DNA molecule comprising a nucleotide sequence encoding a plant 4-coumarate Co-enzyme A ligase operably linked to a promoter so that the

recombinant DNA molecule is expressed in an amount effective to alter the lignin structure of the plant.

45. An expression cassette comprising a transcriptional control region of a 4-coumarate Co-enzyme A ligase gene operably linked to a DNA segment comprising an open reading frame.

46. A method of expressing a DNA segment in the xylem of a plant, comprising:

(a) introducing an expression cassette comprising a transcriptional control region of a 4-coumarate Co-enzyme A ligase gene operably linked to a DNA segment into cells of a plant;

(b) regenerating the plant cells to provide a transgenic plant; and

(c) expressing the DNA segment in the xylem of a plant.

47. A method of expressing a DNA segment in the epidermal tissue of a plant, comprising:

(a) introducing an expression cassette comprising a transcriptional control region of a 4-coumarate Co-enzyme A ligase gene purple operably linked to a DNA segment into cells of a plant;

(b) regenerating the plant cells to provide a transgenic plant; and

(c) expressing the DNA segment in the epidermal tissue of a plant.

48. The method of claim 46 wherein the transgenic plant has altered lignin content, lignin structure, cellulose content or wood quality relative to the corresponding non-transgenic plant.

49. The plant of claim 38 which has altered levels of phenylpropanoids or other secondary metabolities relative to the corresponding non-transgenic plant.

50. The method of claim 1 wherein the plant has enhanced root growth.

51. The plant of claim 10 wherein the plant has enhanced root growth.

52. The method of claim 1 wherein the plant has enhanced root development.

53. The plant of claim 10 wherein the plant has enhanced root development.

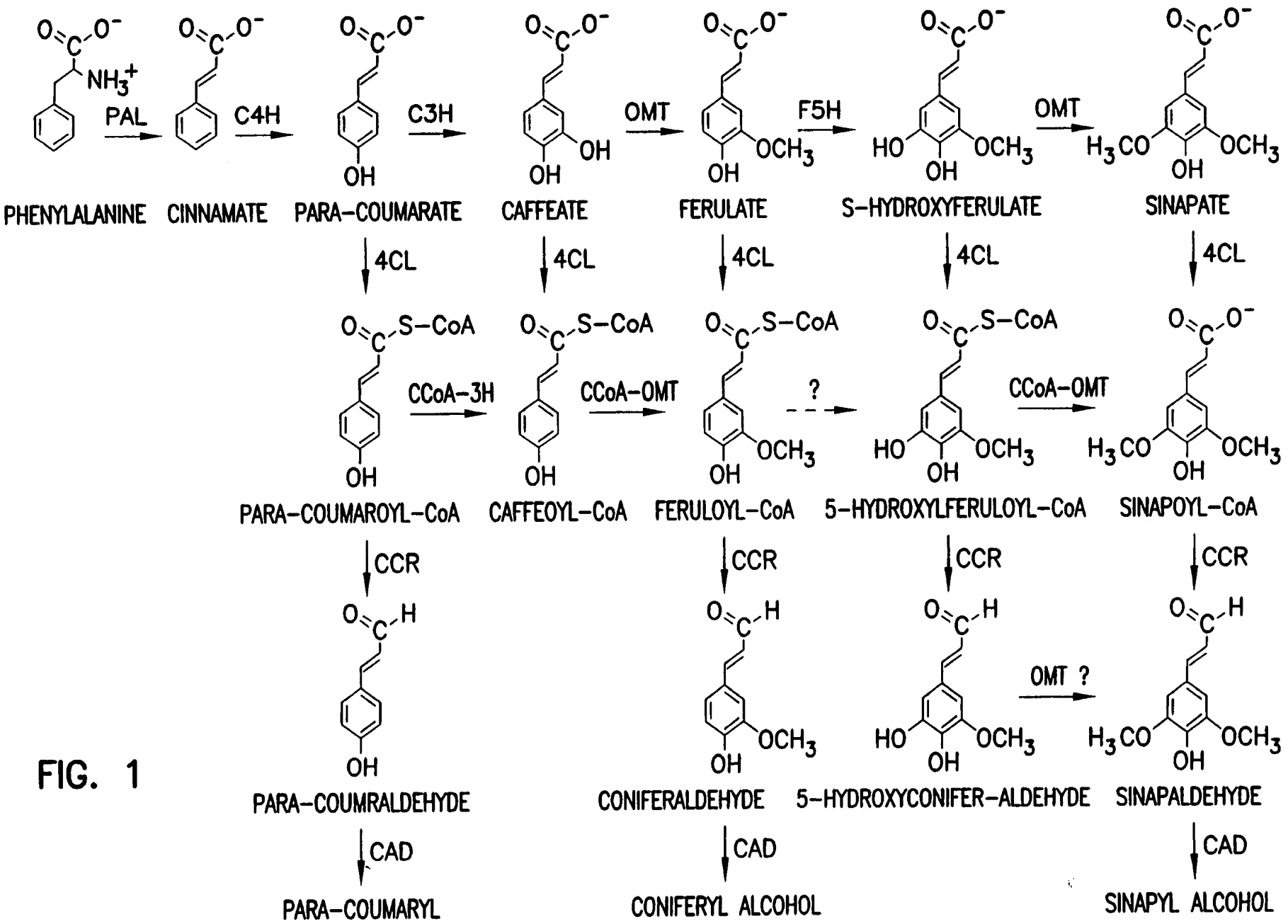


FIG. 1

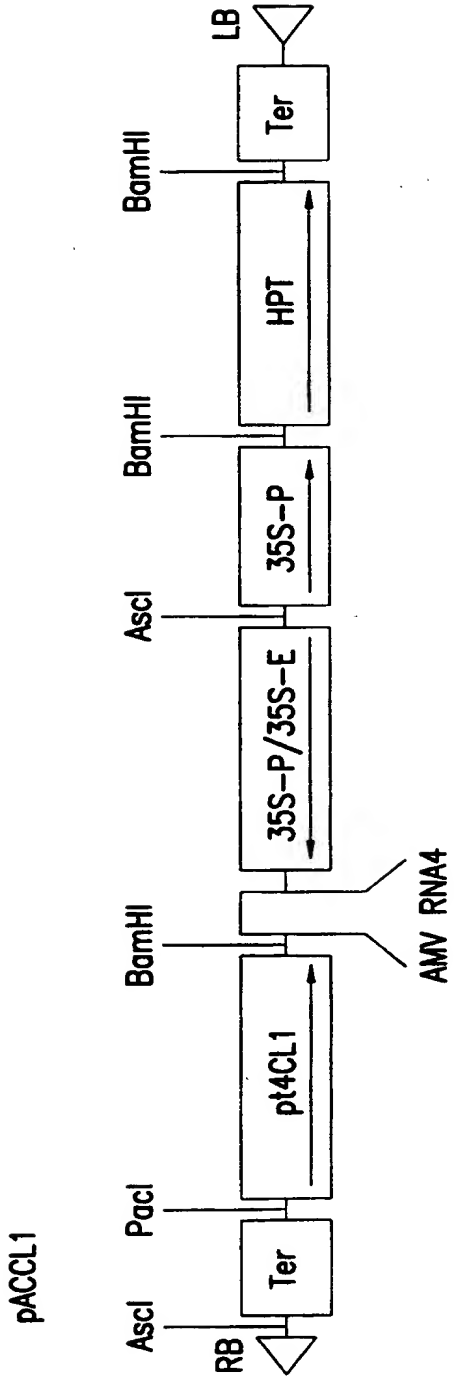

LB
Ter
BamHI
HPT
BamHI
35S-P
AscI
35S-P/35S-E
BamHI
AMV RNA4
pt4CL1
PacI
Ter
AscI
RB
pACCL1

FIG. 2

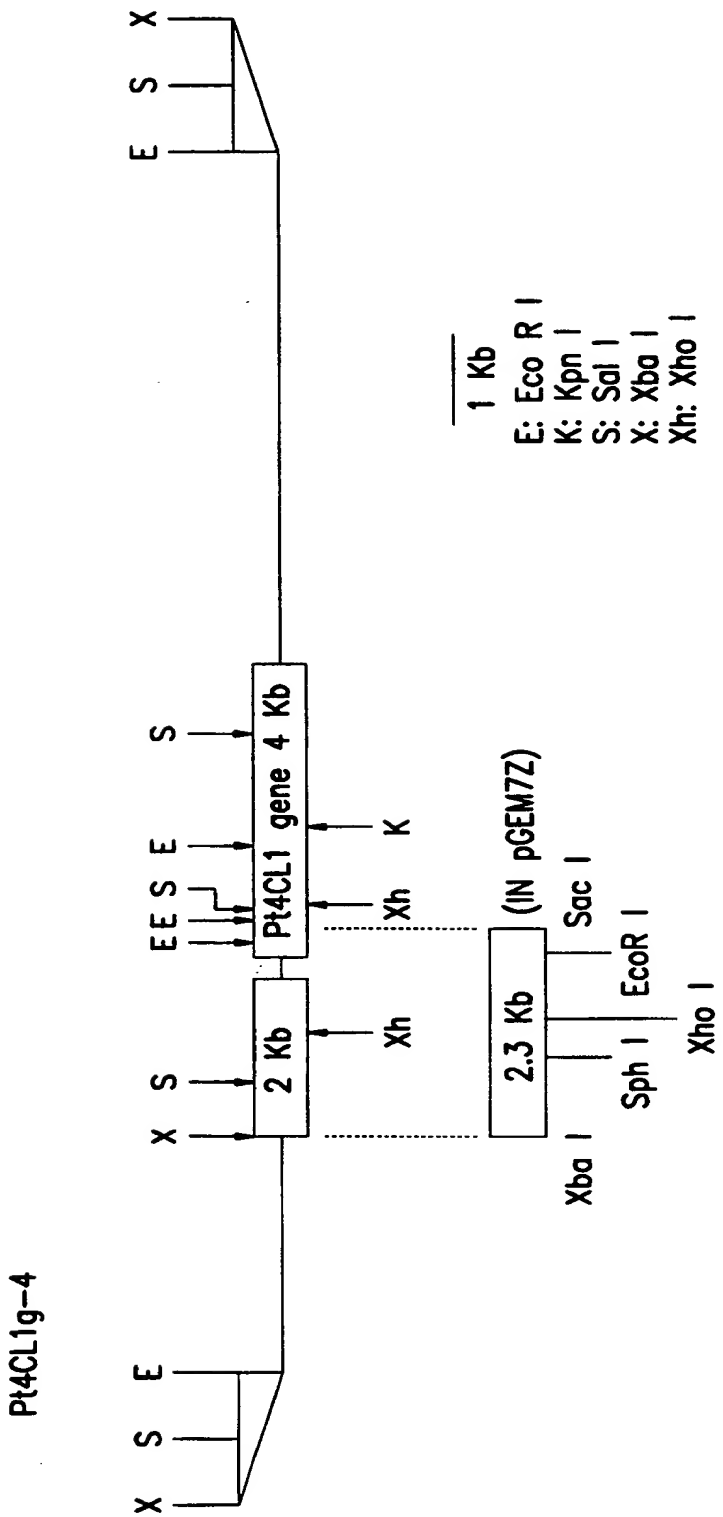

Pt4CL1g-4
X S E
X S E E S S
Pt4CL1 gene 4 Kb
2 Kb
Xh
K
Xh
Xh
(IN pGEM7Z)
2.3 Kb
Sac I
EcoR I
Xho I
Sph I
Xba I
1 Kb
E: Eco R I
K: Kpn I
S: Sal I
X: Xba I
Xh: Xho I

FIG. 3

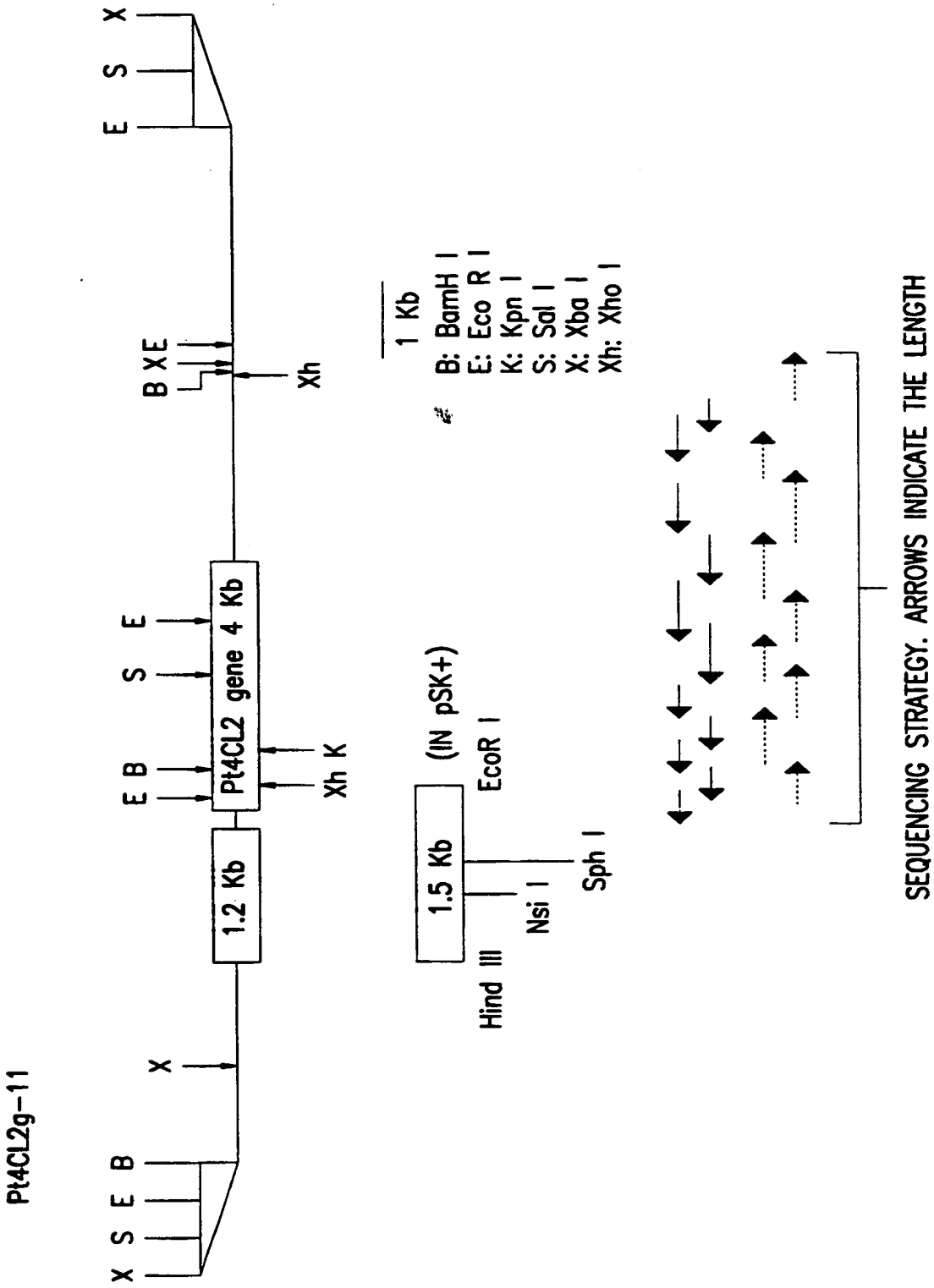

Pt4CL2g-11
X S E B
X
1.2 Kb
E B
S E
Pt4CL2 gene 4 Kb
Xh K
B X E
Xh
E S X
1.5 Kb
(IN pSK+)
EcoR I
Hind III
Nsi I
Sph I
1 Kb
B: BamH I
E: Eco R I
K: Kpn I
S: Sal I
X: Xba I
Xh: Xho I
SEQUENCING STRATEGY. ARROWS INDICATE THE LENGTH AND DIRECTION OF DNA STRANDS SEQUENCED.

FIG. 4

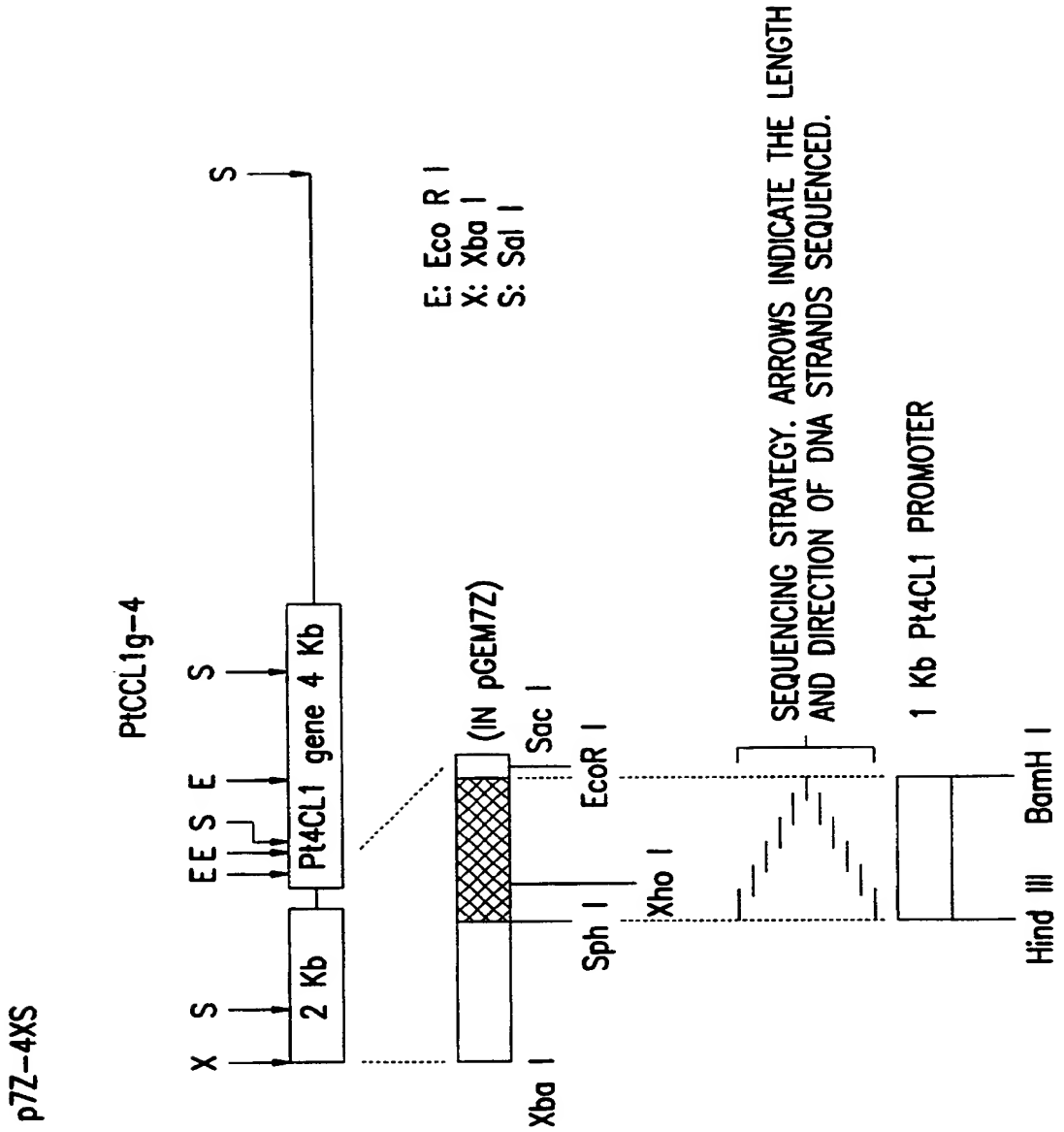

p7Z-4XS
PtCCL1g-4
X S
2 Kb
E E S E S
Pt4CL1 gene 4 Kb
S
E: Eco R I
X: Xba I
S: Sal I
(IN pGEM7Z)
Xba I
Sph I
Xho I
EcoR I
Sac I
SEQUENCING STRATEGY. ARROWS INDICATE THE LENGTH AND DIRECTION OF DNA STRANDS SEQUENCED.
1 Kb Pt4CL1 PROMOTER
Hind III
BamH I

FIG. 5

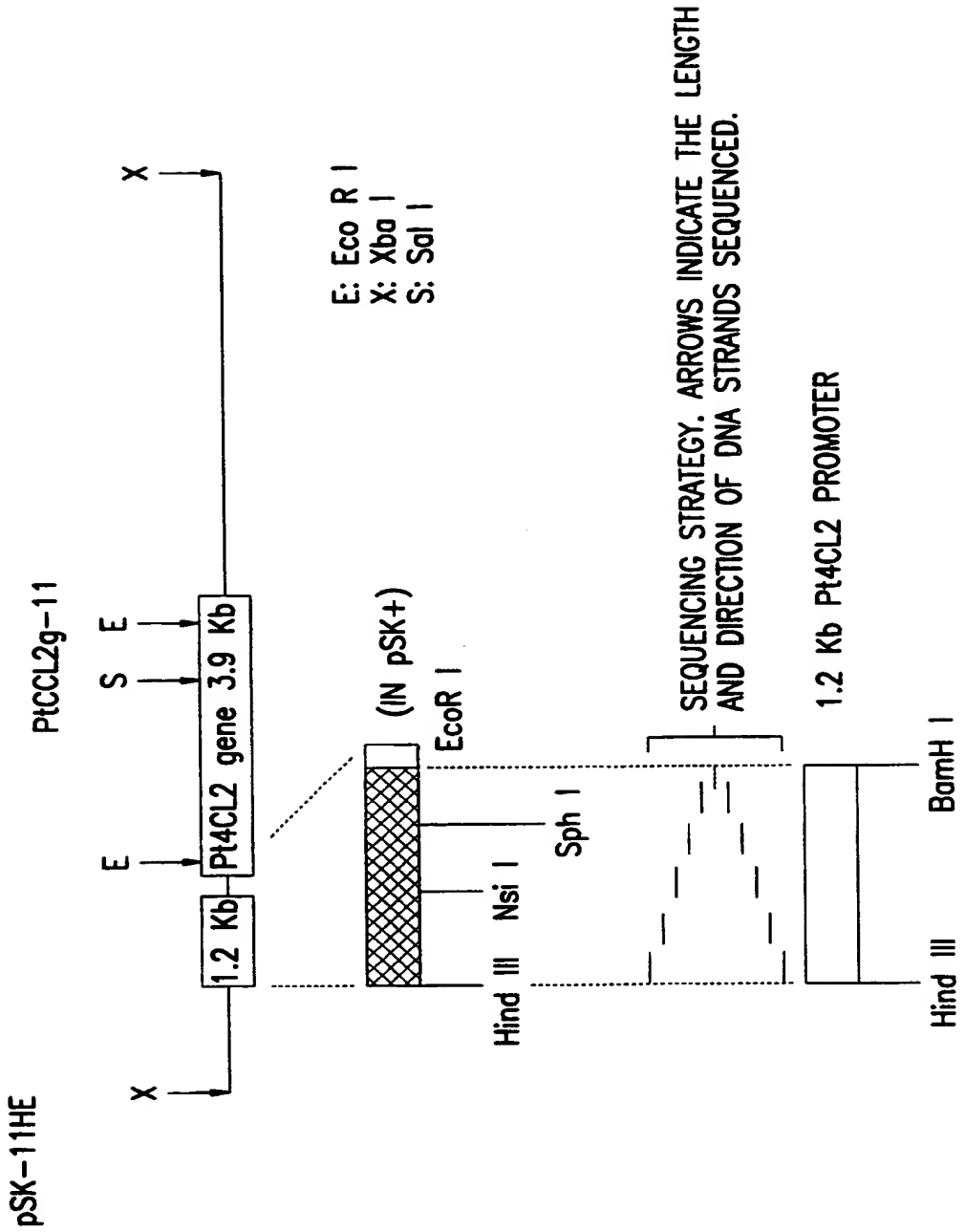

pSK-11HE
PtCCL2g-11
X
S E
Pt4CL2 gene 3.9 Kb
E
1.2 Kb
X
E: Eco R I
X: Xba I
S: Sal I
(IN pSK+)
EcoR I
Sph I
Nsi I
Hind III
SEQUENCING STRATEGY. ARROWS INDICATE THE LENGTH AND DIRECTION OF DNA STRANDS SEQUENCED.
1.2 Kb Pt4CL2 PROMOTER
BamH I
Hind III

FIG. 6

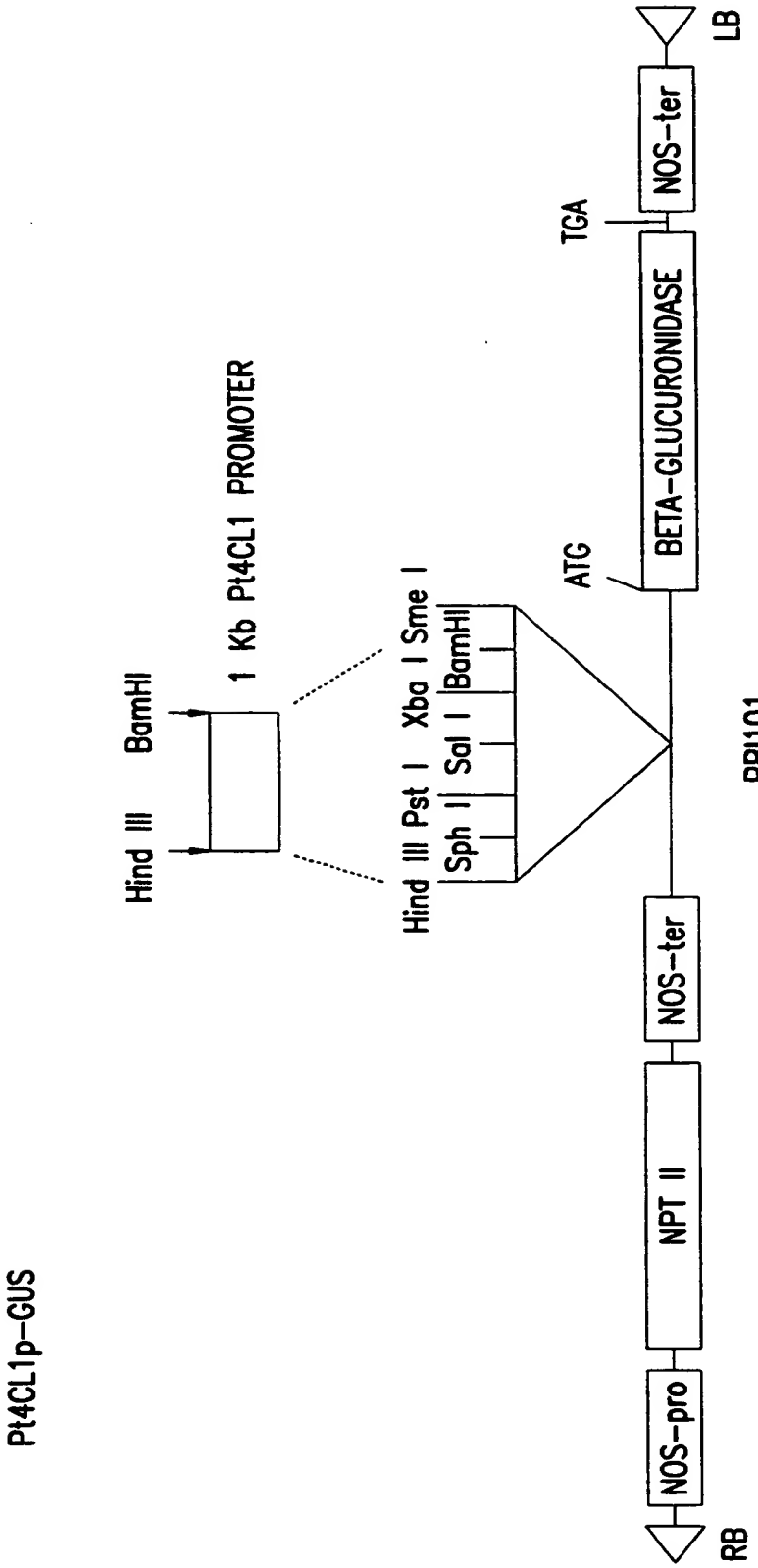

Pt4CL1p-GUS
Hind III
BamHI
1 Kb Pt4CL1 PROMOTER
Hind III Pst I Xba I Sme I
Sph I Sal I BamHI
ATG
BETA-GLUCURONIDASE
TGA
NOS-ter
LB
PBI101
NOS-ter
NPT II
NOS-pro
RB

FIG. 7

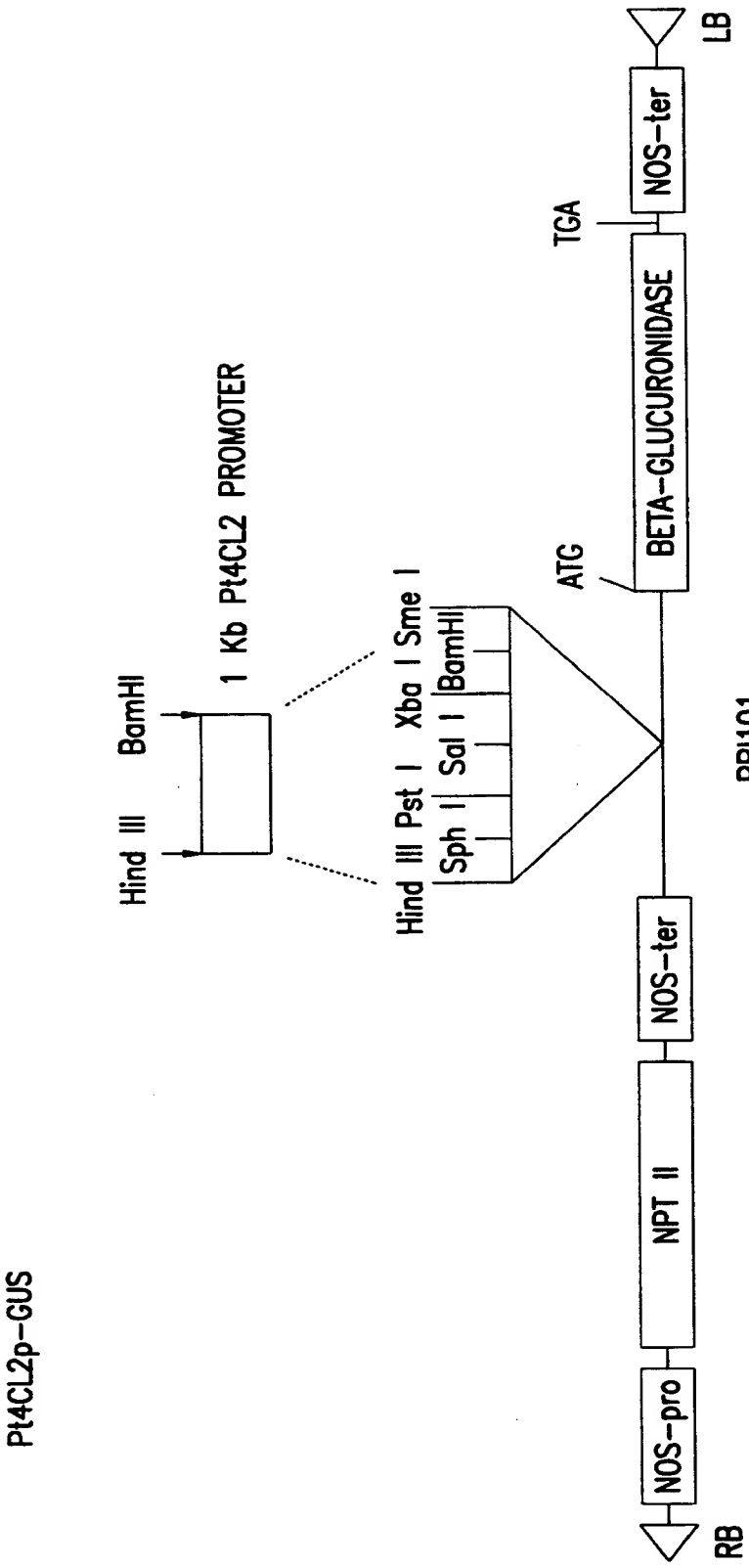

LB
NOS-ter
TGA
BETA-GLUCURONIDASE
ATG
1 Kb Pt4CL2 PROMOTER
Hind III BamHI
Hind III Pst I Xba I Sme I
Sph I Sal I BamHI
NOS-ter
NPT II
NOS-pro
RB
PBI101
Pt4CL2p-GUS

FIG. 8

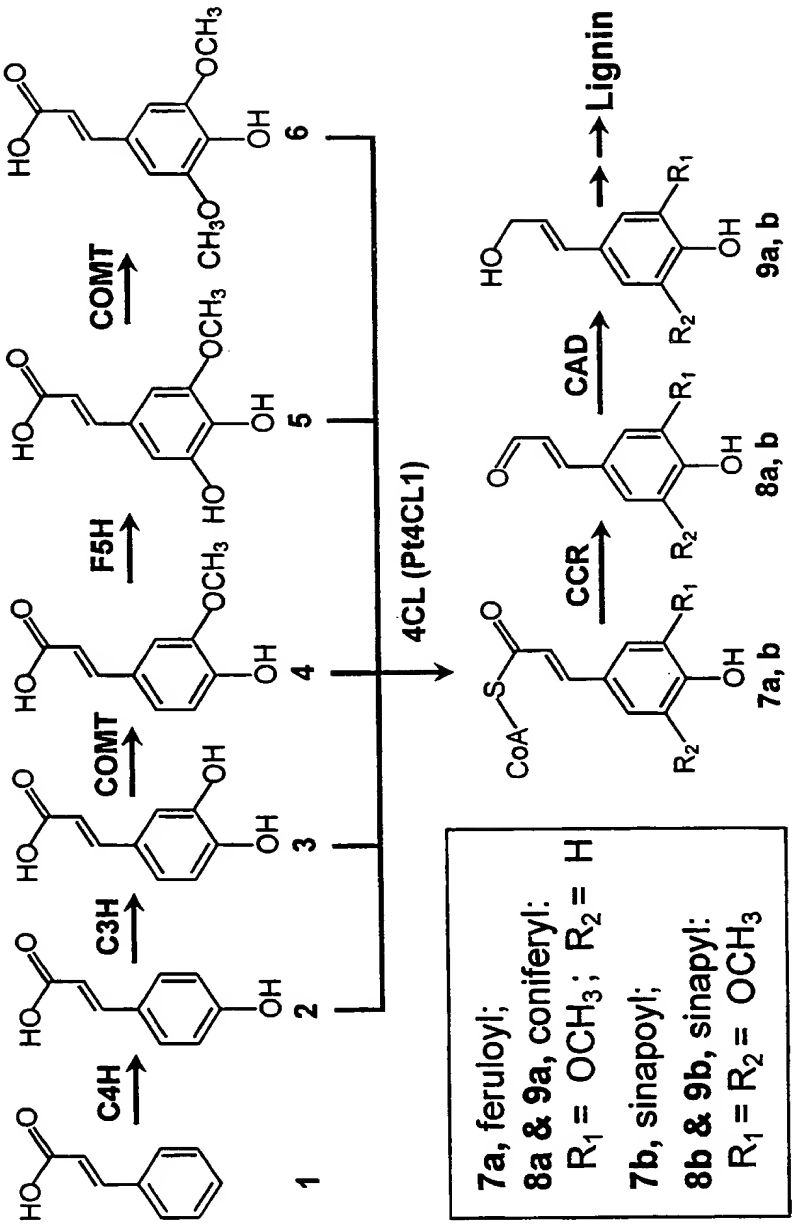

C4H
C3H
COMT
F5H
COMT
4CL (Pt4CL1)
CCR
CAD
Lignin
1
2
3
4
5
6
7a, b
8a, b
9a, b
7a, feruloyl;
8a & 9a, coniferyl:
R1 = OCH3; R2 = H
7b, sinapoyl;
8b & 9b, sinapyl:
R1 = R2 = OCH3

FIG. 9

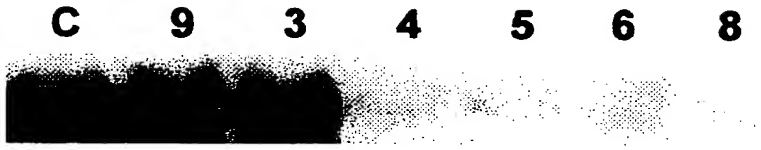


FIG. 10A

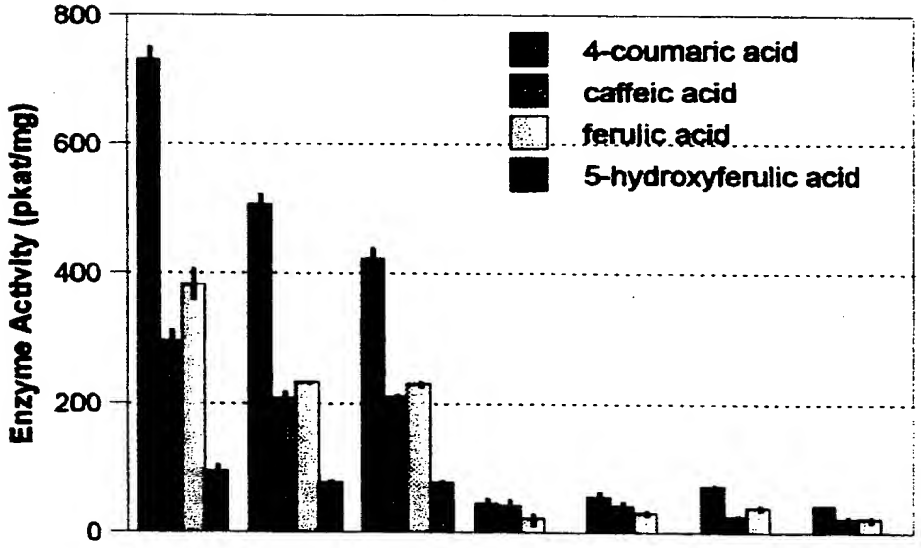


FIG. 10B

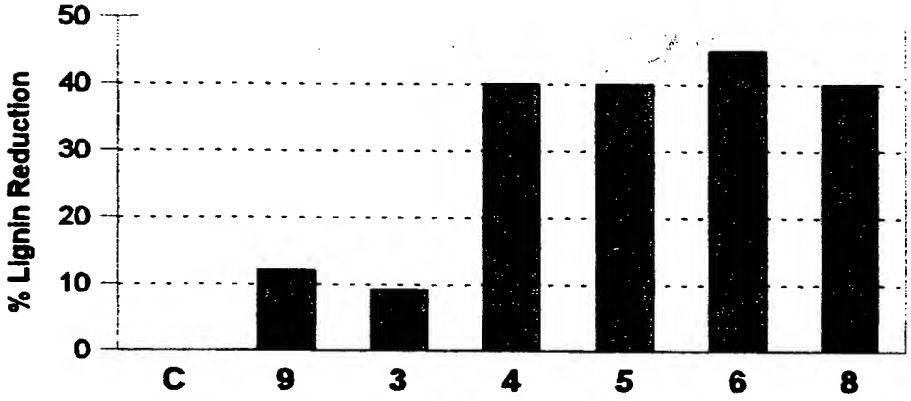


FIG. 10C

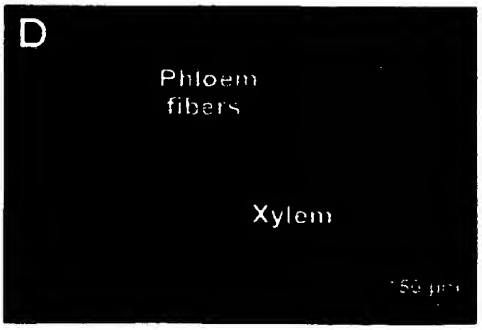


FIG. 10D

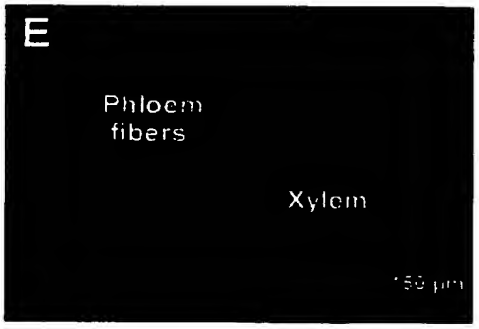


FIG. 10E

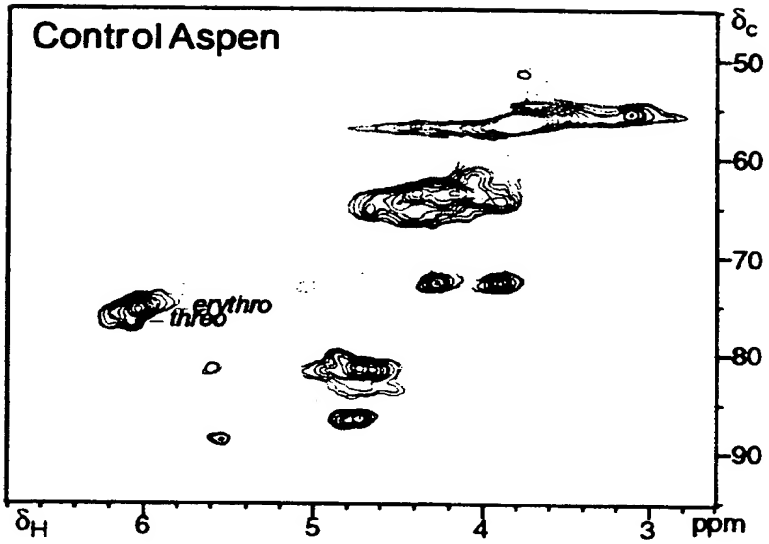


FIG. 11A

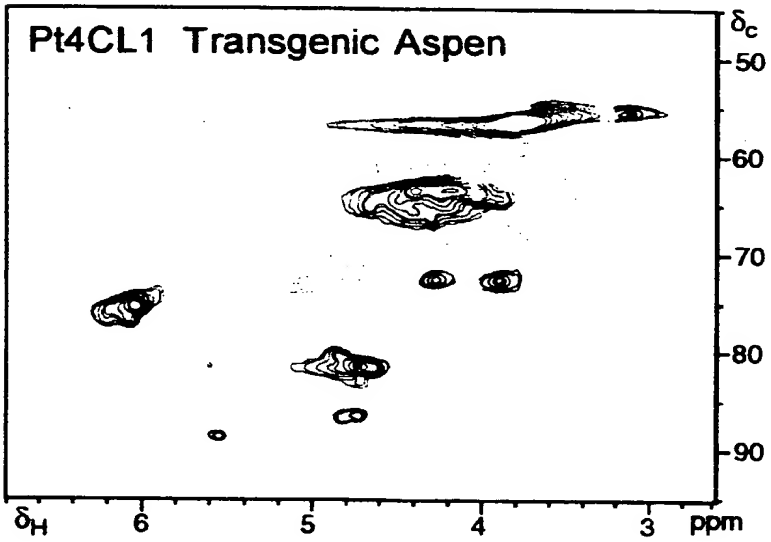


FIG. 11B

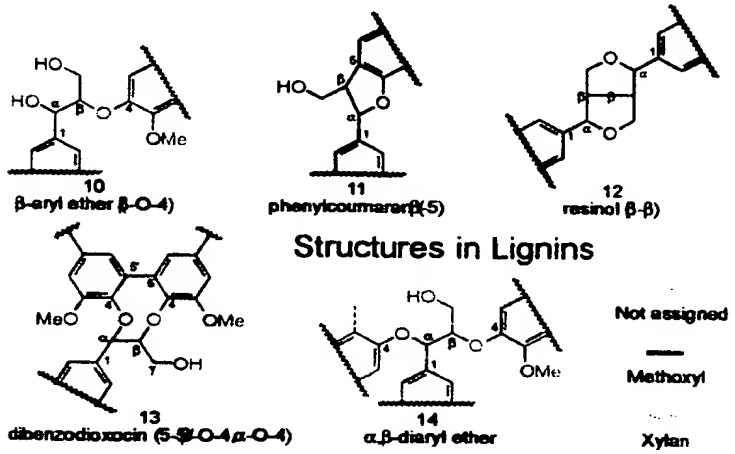


FIG. 11C

FIG. 12A

FIG. 12B

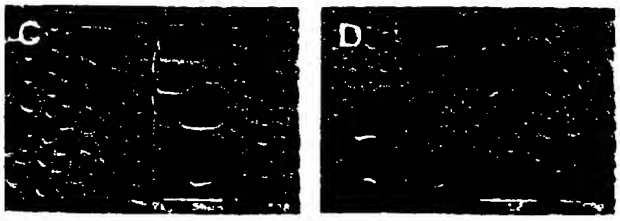


FIG. 12C FIG. 12D

FIG. 12E FIG. 12F

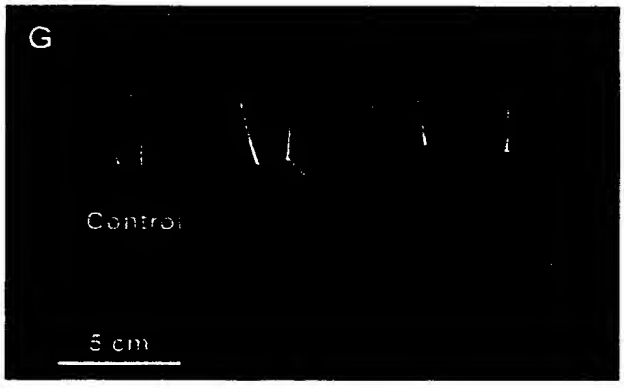


FIG. 12G

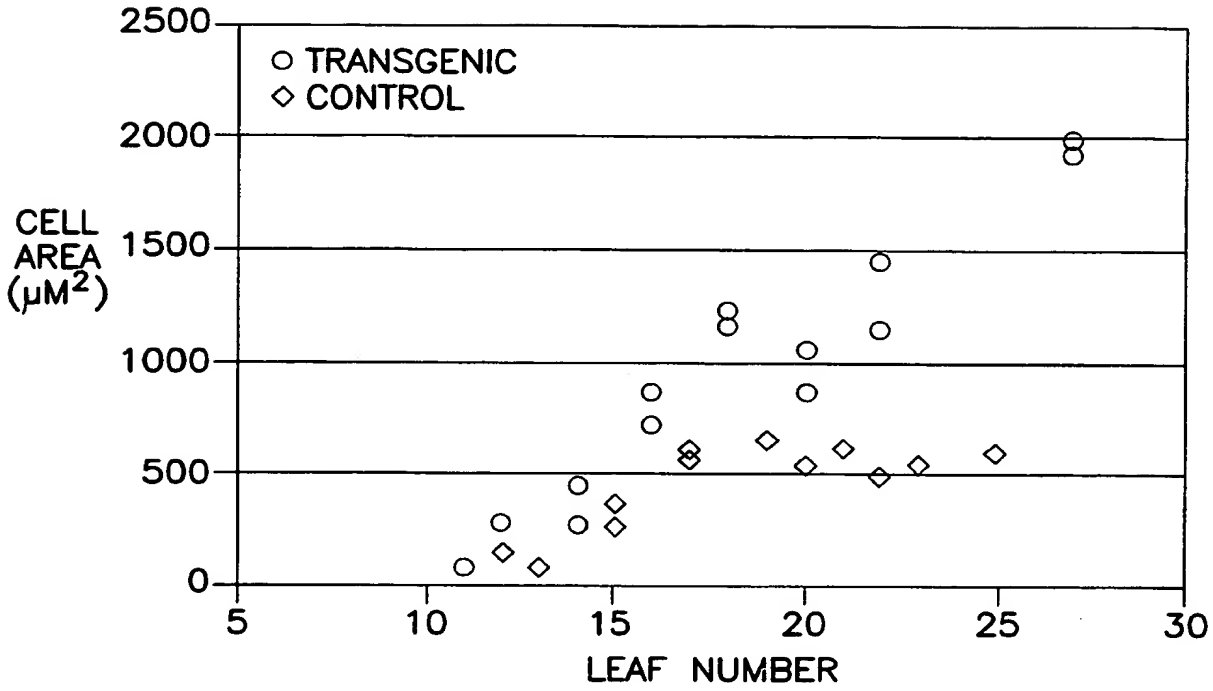

2500
2000
1500
1000
500
0
CELL AREA (μM^2)
TRANSGENIC
CONTROL
5
10
15
20
25
30
LEAF NUMBER

FIG. 12H

SEQUENCE LISTING

<110> The Board of Control of
Michigan Technological University et al.

<120> GENETIC ENGINEERING OF LIGNIN BIOSYNTHESIS IN PLANTS

<130> 881.003WO1

<140>
<141> Herewith

<150> 08/969,046
<151> 1997-11-12

<160> 17

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 1915
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<220>
<221> CDS
<222> (83)...(1687)

<400> 1

```
ccctcgcgaa actccgaaaa cagagagcac ctaaaactca ccatctctcc ctctgcatct      60
ttagcccgca atggacgcca ca atg aat cca caa gaa ttc atc ttt cgc tca      112
                         Met Asn Pro Gln Glu Phe Ile Phe Arg Ser
                          1               5                  10

aaa tta cca gac atc tac atc ccg aaa aac ctt ccc ctg cat tca tac      160
Lys Leu Pro Asp Ile Tyr Ile Pro Lys Asn Leu Pro Leu His Ser Tyr
                 15                  20                  25

gtt ctt gag aac ttg tct aaa cat tca tca aaa cct tgc ctg ata aat      208
Val Leu Glu Asn Leu Ser Lys His Ser Ser Lys Pro Cys Leu Ile Asn
             30                  35                  40

ggc gcg aat gga gat gtc tac acc tat gct gat gtt gag ctc aca gca      256
Gly Ala Asn Gly Asp Val Tyr Thr Tyr Ala Asp Val Glu Leu Thr Ala
         45                  50                  55

aga aga gtt gct tct ggt ctg aac aag att ggt att caa caa ggt gac      304
Arg Arg Val Ala Ser Gly Leu Asn Lys Ile Gly Ile Gln Gln Gly Asp
     60                  65                  70

gtg atc atg ctc ttc cta cca agt tca cct gaa ttc gtg ctt gct ttc      352
Val Ile Met Leu Phe Leu Pro Ser Ser Pro Glu Phe Val Leu Ala Phe
 75                  80                  85                  90

cta ggc gct tca cac aga ggt gcc atg atc act gct gcc aat cct ttc      400
Leu Gly Ala Ser His Arg Gly Ala Met Ile Thr Ala Ala Asn Pro Phe
                 95                 100                 105

tcc acc cct gca gag cta gca aaa cat gcc aag gcc tcg aga gca aag      448
Ser Thr Pro Ala Glu Leu Ala Lys His Ala Lys Ala Ser Arg Ala Lys
```

```
                110                 115                 120

ctt ctg ata aca cag gct tgt tac tac gag aag gtt aaa gat ttt gcc      496
Leu Leu Ile Thr Gln Ala Cys Tyr Tyr Glu Lys Val Lys Asp Phe Ala
        125                 130                 135

cga gaa agt gat gtt aag gtc atg tgc gtg gac tct gcc ccg gac ggt      544
Arg Glu Ser Asp Val Lys Val Met Cys Val Asp Ser Ala Pro Asp Gly
    140                 145                 150

gct tca ctt ttc aga gct cac aca cag gca gac gaa aat gaa gtg cct      592
Ala Ser Leu Phe Arg Ala His Thr Gln Ala Asp Glu Asn Glu Val Pro
155                 160                 165                 170

cag gtc gac att agt cct gat gat gtc gta gca ttg cct tat tca tca      640
Gln Val Asp Ile Ser Pro Asp Asp Val Val Ala Leu Pro Tyr Ser Ser
                175                 180                 185

ggg act aca ggg ttg cca aaa ggg gtc atg tta acg cac aaa ggg cta      688
Gly Thr Thr Gly Leu Pro Lys Gly Val Met Leu Thr His Lys Gly Leu
            190                 195                 200

ata acc agt gtg gct caa cag gta gat gga gac aat cct aac ctg tat      736
Ile Thr Ser Val Ala Gln Gln Val Asp Gly Asp Asn Pro Asn Leu Tyr
        205                 210                 215

ttt cac agt gaa gat gtg att ctg tgt gtg ctt cct atg ttc cat atc      784
Phe His Ser Glu Asp Val Ile Leu Cys Val Leu Pro Met Phe His Ile
    220                 225                 230

tat gct ctg aat tca atg atg ctc tgt ggt ctg aga gtt ggt gcc tcg      832
Tyr Ala Leu Asn Ser Met Met Leu Cys Gly Leu Arg Val Gly Ala Ser
235                 240                 245                 250

att ttg ata atg cca aag ttt gag att ggt tct ttg ctg gga ttg att      880
Ile Leu Ile Met Pro Lys Phe Glu Ile Gly Ser Leu Leu Gly Leu Ile
                255                 260                 265

gag aag tac aag gta tct ata gca cca gtt gtt cca cct gtg atg atg      928
Glu Lys Tyr Lys Val Ser Ile Ala Pro Val Val Pro Pro Val Met Met
            270                 275                 280

gca att gct aag tca cct gat ctt gac aag cat gac ctg tct tct ttg      976
Ala Ile Ala Lys Ser Pro Asp Leu Asp Lys His Asp Leu Ser Ser Leu
        285                 290                 295

agg atg ata aaa tct gga ggg gct cca ttg ggc aag gaa ctt gaa gat     1024
Arg Met Ile Lys Ser Gly Gly Ala Pro Leu Gly Lys Glu Leu Glu Asp
    300                 305                 310

act gtc aga gct aag ttt cct cag gct aga ctt ggt cag gga tat gga     1072
Thr Val Arg Ala Lys Phe Pro Gln Ala Arg Leu Gly Gln Gly Tyr Gly
315                 320                 325                 330

atg acc gag gca gga cct gtt cta gca atg tgc ttg gca ttt gcc aag     1120
Met Thr Glu Ala Gly Pro Val Leu Ala Met Cys Leu Ala Phe Ala Lys
                335                 340                 345

gaa cca ttc gac ata aaa cca ggt gca tgt gga act gta gtc agg aat     1168
Glu Pro Phe Asp Ile Lys Pro Gly Ala Cys Gly Thr Val Val Arg Asn
            350                 355                 360
```

```
gca gag atg aag att gtt gac cca gaa aca ggg gtc tct cta ccg agg      1216
Ala Glu Met Lys Ile Val Asp Pro Glu Thr Gly Val Ser Leu Pro Arg
        365                 370                 375

aac cag cct ggt gag atc tgc atc cgg ggt gat cag atc atg aaa gga      1264
Asn Gln Pro Gly Glu Ile Cys Ile Arg Gly Asp Gln Ile Met Lys Gly
    380                 385                 390

tat ctt aat gac ccc gag gca acc tca aga aca ata gac aaa gaa gga      1312
Tyr Leu Asn Asp Pro Glu Ala Thr Ser Arg Thr Ile Asp Lys Glu Gly
395                 400                 405                 410

tgg ctg cac aca ggc gat atc ggc tac att gat gat gat gat gag ctt      1360
Trp Leu His Thr Gly Asp Ile Gly Tyr Ile Asp Asp Asp Asp Glu Leu
                415                 420                 425

ttc atc gtt gac aga ttg aag gaa ttg atc aag tat aaa ggg ttt cag      1408
Phe Ile Val Asp Arg Leu Lys Glu Leu Ile Lys Tyr Lys Gly Phe Gln
            430                 435                 440

gtt gct cct act gaa ctc gaa gct ttg tta ata gcc cat cca gag ata      1456
Val Ala Pro Thr Glu Leu Glu Ala Leu Leu Ile Ala His Pro Glu Ile
        445                 450                 455

tcc gat gct gct gta gta gga ttg aaa gat gag gat gcg gga gaa gtt      1504
Ser Asp Ala Ala Val Val Gly Leu Lys Asp Glu Asp Ala Gly Glu Val
    460                 465                 470

cct gtt gca ttt gta gtg aaa tca gaa aag tct cag gcc acc gaa gat      1552
Pro Val Ala Phe Val Val Lys Ser Glu Lys Ser Gln Ala Thr Glu Asp
475                 480                 485                 490

gaa att aag cag tat att tca aaa cag gtg atc ttc tac aag aga ata      1600
Glu Ile Lys Gln Tyr Ile Ser Lys Gln Val Ile Phe Tyr Lys Arg Ile
                495                 500                 505

aaa cga gtt ttc ttc att gaa gca att ccc aag gca cca tca ggc aag      1648
Lys Arg Val Phe Phe Ile Glu Ala Ile Pro Lys Ala Pro Ser Gly Lys
            510                 515                 520

atc ctg agg aag aat ctg aaa gag aag ttg cca ggc ata taactgaaga      1697
Ile Leu Arg Lys Asn Leu Lys Glu Lys Leu Pro Gly Ile
        525                 530                 535

tgttactgaa catttaaccc tctgtcttat ttctttaata cttgcgaatc attgtagtgt   1757
tgaaccaagc atgcttggaa aagacacgta cccaacgtaa gacagttact gttcctagta   1817
tacaagctct ttaatgttcg ttttgaactt gggaaaacat aagttctcct gtcgccatat   1877
ggagtaattc aattgaatat tttggtttct ttaatgat                           1915
```

<210> 2
<211> 535
<212> PRT
<213> Populus tremuloides Michx. (aspen)

<400> 2

```
Met Asn Pro Gln Glu Phe Ile Phe Arg Ser Lys Leu Pro Asp Ile Tyr
 1               5                  10                  15
Ile Pro Lys Asn Leu Pro Leu His Ser Tyr Val Leu Glu Asn Leu Ser
            20                  25                  30
```

Lys His Ser Ser Lys Pro Cys Leu Ile Asn Gly Ala Asn Gly Asp Val
35 40 45
Tyr Thr Tyr Ala Asp Val Glu Leu Thr Ala Arg Arg Val Ala Ser Gly
50 55 60
Leu Asn Lys Ile Gly Ile Gln Gln Gly Asp Val Ile Met Leu Phe Leu
65 70 75 80
Pro Ser Ser Pro Glu Phe Val Leu Ala Phe Leu Gly Ala Ser His Arg
85 90 95
Gly Ala Met Ile Thr Ala Ala Asn Pro Phe Ser Thr Pro Ala Glu Leu
100 105 110
Ala Lys His Ala Lys Ala Ser Arg Ala Lys Leu Leu Ile Thr Gln Ala
115 120 125
Cys Tyr Tyr Glu Lys Val Lys Asp Phe Ala Arg Glu Ser Asp Val Lys
130 135 140
Val Met Cys Val Asp Ser Ala Pro Asp Gly Ala Ser Leu Phe Arg Ala
145 150 155 160
His Thr Gln Ala Asp Glu Asn Glu Val Pro Gln Val Asp Ile Ser Pro
165 170 175
Asp Asp Val Val Ala Leu Pro Tyr Ser Ser Gly Thr Thr Gly Leu Pro
180 185 190
Lys Gly Val Met Leu Thr His Lys Gly Leu Ile Thr Ser Val Ala Gln
195 200 205
Gln Val Asp Gly Asp Asn Pro Asn Leu Tyr Phe His Ser Glu Asp Val
210 215 220
Ile Leu Cys Val Leu Pro Met Phe His Ile Tyr Ala Leu Asn Ser Met
225 230 235 240
Met Leu Cys Gly Leu Arg Val Gly Ala Ser Ile Leu Ile Met Pro Lys
245 250 255
Phe Glu Ile Gly Ser Leu Leu Gly Leu Ile Glu Lys Tyr Lys Val Ser
260 265 270
Ile Ala Pro Val Val Pro Pro Val Met Met Ala Ile Ala Lys Ser Pro
275 280 285
Asp Leu Asp Lys His Asp Leu Ser Ser Leu Arg Met Ile Lys Ser Gly
290 295 300
Gly Ala Pro Leu Gly Lys Glu Leu Glu Asp Thr Val Arg Ala Lys Phe
305 310 315 320
Pro Gln Ala Arg Leu Gly Gln Gly Tyr Gly Met Thr Glu Ala Gly Pro
325 330 335
Val Leu Ala Met Cys Leu Ala Phe Ala Lys Glu Pro Phe Asp Ile Lys
340 345 350
Pro Gly Ala Cys Gly Thr Val Val Arg Asn Ala Glu Met Lys Ile Val
355 360 365
Asp Pro Glu Thr Gly Val Ser Leu Pro Arg Asn Gln Pro Gly Glu Ile
370 375 380
Cys Ile Arg Gly Asp Gln Ile Met Lys Gly Tyr Leu Asn Asp Pro Glu
385 390 395 400
Ala Thr Ser Arg Thr Ile Asp Lys Glu Gly Trp Leu His Thr Gly Asp
405 410 415
Ile Gly Tyr Ile Asp Asp Asp Asp Glu Leu Phe Ile Val Asp Arg Leu
420 425 430
Lys Glu Leu Ile Lys Tyr Lys Gly Phe Gln Val Ala Pro Thr Glu Leu
435 440 445
Glu Ala Leu Leu Ile Ala His Pro Glu Ile Ser Asp Ala Ala Val Val
450 455 460
Gly Leu Lys Asp Glu Asp Ala Gly Glu Val Pro Val Ala Phe Val Val
465 470 475 480
Lys Ser Glu Lys Ser Gln Ala Thr Glu Asp Glu Ile Lys Gln Tyr Ile
485 490 495
Ser Lys Gln Val Ile Phe Tyr Lys Arg Ile Lys Arg Val Phe Phe Ile
500 505 510
Glu Ala Ile Pro Lys Ala Pro Ser Gly Lys Ile Leu Arg Lys Asn Leu

```
        515                 520                 525
Lys Glu Lys Leu Pro Gly Ile
    530                 535
```

<210> 3
<211> 1710
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<220>
<221> CDS
<222> (1)...(1710)

<400> 3

```
atg atg tcc gtg gcc acg gtt gag ccc ccg aaa ccg gaa ctc tcc cct        48
Met Met Ser Val Ala Thr Val Glu Pro Pro Lys Pro Glu Leu Ser Pro
 1               5                  10                  15

cca caa aac caa aac gca cca tcc tct cat gaa act gat cac att ttc        96
Pro Gln Asn Gln Asn Ala Pro Ser Ser His Glu Thr Asp His Ile Phe
            20                  25                  30

aga tca aaa cta cca gac ata acc atc tcg aac gac ctc cct ctg cac       144
Arg Ser Lys Leu Pro Asp Ile Thr Ile Ser Asn Asp Leu Pro Leu His
        35                  40                  45

gca tac tgc ttt gaa aac ctc tct gat ttc tca gat agg cca tgc ttg       192
Ala Tyr Cys Phe Glu Asn Leu Ser Asp Phe Ser Asp Arg Pro Cys Leu
    50                  55                  60

att tca ggt tcc acg gga aaa acc tat tct ttt gcc gaa act cac ctc       240
Ile Ser Gly Ser Thr Gly Lys Thr Tyr Ser Phe Ala Glu Thr His Leu
65                  70                  75                  80

ata tct cgg aag gtc gct gct ggg tta tcc aat ttg ggc atc aag aaa       288
Ile Ser Arg Lys Val Ala Ala Gly Leu Ser Asn Leu Gly Ile Lys Lys
                85                  90                  95

ggc gat gta atc atg acc ctg ctc caa aac tgc cca gaa ttc gtc ttc       336
Gly Asp Val Ile Met Thr Leu Leu Gln Asn Cys Pro Glu Phe Val Phe
            100                 105                 110

tcc ttc atc ggt gct tcc atg att ggt gca gtc atc acc act gcg aac       384
Ser Phe Ile Gly Ala Ser Met Ile Gly Ala Val Ile Thr Thr Ala Asn
        115                 120                 125

cct ttc tac act caa agt gaa ata ttc aag caa ttc tct gct tct cgt       432
Pro Phe Tyr Thr Gln Ser Glu Ile Phe Lys Gln Phe Ser Ala Ser Arg
    130                 135                 140

gcg aaa ctg att atc acc cag tct caa tat gtg aac aag cta gga gat       480
Ala Lys Leu Ile Ile Thr Gln Ser Gln Tyr Val Asn Lys Leu Gly Asp
145                 150                 155                 160

agt gat tgc cat gaa aac aac caa aaa ccg ggg gaa gat ttc ata gta       528
Ser Asp Cys His Glu Asn Asn Gln Lys Pro Gly Glu Asp Phe Ile Val
                165                 170                 175

atc acc att gat gac ccg cca gag aac tgt cta cat ttc aat gtg ctt       576
Ile Thr Ile Asp Asp Pro Pro Glu Asn Cys Leu His Phe Asn Val Leu
```

```
                180                 185                 190

gtc gag gct agc gag agt gaa atg cca aca gtt tca atc ctt ccg gat      624
Val Glu Ala Ser Glu Ser Glu Met Pro Thr Val Ser Ile Leu Pro Asp
        195                 200                 205

gat cct gtg gca tta cca ttc tct tca ggg aca aca ggg ctc cca aaa      672
Asp Pro Val Ala Leu Pro Phe Ser Ser Gly Thr Thr Gly Leu Pro Lys
    210                 215                 220

gga gtg ata ctg acc cac aag agc ttg ata aca agt gtg gct caa caa      720
Gly Val Ile Leu Thr His Lys Ser Leu Ile Thr Ser Val Ala Gln Gln
225                 230                 235                 240

gtt gat gga gag atc cca aat tta tac ttg aaa caa gat gac gtt gtt      768
Val Asp Gly Glu Ile Pro Asn Leu Tyr Leu Lys Gln Asp Asp Val Val
                245                 250                 255

tta tgc gtt tta cct ttg ttt cac atc ttt tca ttg aac agc gtg ttg      816
Leu Cys Val Leu Pro Leu Phe His Ile Phe Ser Leu Asn Ser Val Leu
            260                 265                 270

tta tgc tcg ttg aga gcc ggt tct gct gtt ctt tta atg caa aag ttt      864
Leu Cys Ser Leu Arg Ala Gly Ser Ala Val Leu Leu Met Gln Lys Phe
        275                 280                 285

gag ata gga tca ctg cta gag ctc att cag aaa cac aat gtt tcg gtt      912
Glu Ile Gly Ser Leu Leu Glu Leu Ile Gln Lys His Asn Val Ser Val
    290                 295                 300

gcg gct gtg gtg cca cca ctg gtg ctg gcg ttg gcc aag aac cca ttg      960
Ala Ala Val Val Pro Pro Leu Val Leu Ala Leu Ala Lys Asn Pro Leu
305                 310                 315                 320

gag gcg aac ttc gac ttg agt tcg atc agg gta gtc ctg tca ggg gct     1008
Glu Ala Asn Phe Asp Leu Ser Ser Ile Arg Val Val Leu Ser Gly Ala
                325                 330                 335

gcg cca ctg ggg aag gag ctc gag gac gcc ctc agg agc agg gtt cct     1056
Ala Pro Leu Gly Lys Glu Leu Glu Asp Ala Leu Arg Ser Arg Val Pro
            340                 345                 350

cag gcc atc ctg gga cag ggt tat ggg atg aca gag gcc ggg cct gtg     1104
Gln Ala Ile Leu Gly Gln Gly Tyr Gly Met Thr Glu Ala Gly Pro Val
        355                 360                 365

cta tca atg tgc tta gcc ttt tca aag caa cct ttc cca acc aag tct     1152
Leu Ser Met Cys Leu Ala Phe Ser Lys Gln Pro Phe Pro Thr Lys Ser
    370                 375                 380

ggg tcg tgt gga acg gtg gtt aga aac gca gag ctc aag gtc att gac     1200
Gly Ser Cys Gly Thr Val Val Arg Asn Ala Glu Leu Lys Val Ile Asp
385                 390                 395                 400

cct gag acc ggt cgc tct ctt ggt tac aac caa cct ggt gaa atc tgc     1248
Pro Glu Thr Gly Arg Ser Leu Gly Tyr Asn Gln Pro Gly Glu Ile Cys
                405                 410                 415

atc cgt gga tcc caa atc atg aaa gga tat ttg aat gac gcg gaa gcc     1296
Ile Arg Gly Ser Gln Ile Met Lys Gly Tyr Leu Asn Asp Ala Glu Ala
            420                 425                 430
```

```
acg gca aac acc ata gac gtt gag ggt tgg ctc cac act gga gat ata      1344
Thr Ala Asn Thr Ile Asp Val Glu Gly Trp Leu His Thr Gly Asp Ile
        435                 440                 445

ggt tat gtc gac gac gac gac gag att ttc att gtt gat aga gtg aag      1392
Gly Tyr Val Asp Asp Asp Asp Glu Ile Phe Ile Val Asp Arg Val Lys
    450                 455                 460

gaa atc ata aaa ttc aaa ggc ttc cag gtg ccg cca gcg gag ctt gag      1440
Glu Ile Ile Lys Phe Lys Gly Phe Gln Val Pro Pro Ala Glu Leu Glu
465                 470                 475                 480

gct ctc ctt gta aac cac cct tca att gcg gat gcg gct gtt gtt ccg      1488
Ala Leu Leu Val Asn His Pro Ser Ile Ala Asp Ala Ala Val Val Pro
                485                 490                 495

caa aaa gac gag gtt gct ggt gaa gtt cct gtc gcg ttt gtg gtc cgc      1536
Gln Lys Asp Glu Val Ala Gly Glu Val Pro Val Ala Phe Val Val Arg
            500                 505                 510

tca gat gat ctt gac ctt agt gaa gag gct gta aaa gaa tac att gca      1584
Ser Asp Asp Leu Asp Leu Ser Glu Glu Ala Val Lys Glu Tyr Ile Ala
        515                 520                 525

aag cag gtg gtg ttc tac aag aaa ctg cac aag gtg ttc ttc gtt cat      1632
Lys Gln Val Val Phe Tyr Lys Lys Leu His Lys Val Phe Phe Val His
    530                 535                 540

tct att ccc aaa tcg gct tct gga aag att cta aga aaa gac ctc aga      1680
Ser Ile Pro Lys Ser Ala Ser Gly Lys Ile Leu Arg Lys Asp Leu Arg
545                 550                 555                 560

gcc aag ctt gcc aca gcc acc acc atg tcc                             1710
Ala Lys Leu Ala Thr Ala Thr Thr Met Ser
                565                 570
```

<210> 4
<211> 570
<212> PRT
<213> Populus tremuloides Michx. (aspen)

<400> 4

```
Met Met Ser Val Ala Thr Val Glu Pro Pro Lys Pro Glu Leu Ser Pro
 1               5                  10                  15
Pro Gln Asn Gln Asn Ala Pro Ser Ser His Glu Thr Asp His Ile Phe
            20                  25                  30
Arg Ser Lys Leu Pro Asp Ile Thr Ile Ser Asn Asp Leu Pro Leu His
        35                  40                  45
Ala Tyr Cys Phe Glu Asn Leu Ser Asp Phe Ser Asp Arg Pro Cys Leu
    50                  55                  60
Ile Ser Gly Ser Thr Gly Lys Thr Tyr Ser Phe Ala Glu Thr His Leu
65                  70                  75                  80
Ile Ser Arg Lys Val Ala Ala Gly Leu Ser Asn Leu Gly Ile Lys Lys
                85                  90                  95
Gly Asp Val Ile Met Thr Leu Leu Gln Asn Cys Pro Glu Phe Val Phe
            100                 105                 110
Ser Phe Ile Gly Ala Ser Met Ile Gly Ala Val Ile Thr Thr Ala Asn
        115                 120                 125
Pro Phe Tyr Thr Gln Ser Glu Ile Phe Lys Gln Phe Ser Ala Ser Arg
```

```
    130                 135                 140
Ala Lys Leu Ile Ile Thr Gln Ser Gln Tyr Val Asn Lys Leu Gly Asp
145                 150                 155                 160
Ser Asp Cys His Glu Asn Asn Gln Lys Pro Gly Glu Asp Phe Ile Val
                165                 170                 175
Ile Thr Ile Asp Asp Pro Pro Glu Asn Cys Leu His Phe Asn Val Leu
            180                 185                 190
Val Glu Ala Ser Glu Ser Glu Met Pro Thr Val Ser Ile Leu Pro Asp
        195                 200                 205
Asp Pro Val Ala Leu Pro Phe Ser Ser Gly Thr Thr Gly Leu Pro Lys
    210                 215                 220
Gly Val Ile Leu Thr His Lys Ser Leu Ile Thr Ser Val Ala Gln Gln
225                 230                 235                 240
Val Asp Gly Glu Ile Pro Asn Leu Tyr Leu Lys Gln Asp Asp Val Val
                245                 250                 255
Leu Cys Val Leu Pro Leu Phe His Ile Phe Ser Leu Asn Ser Val Leu
            260                 265                 270
Leu Cys Ser Leu Arg Ala Gly Ser Ala Val Leu Leu Met Gln Lys Phe
        275                 280                 285
Glu Ile Gly Ser Leu Leu Glu Leu Ile Gln Lys His Asn Val Ser Val
    290                 295                 300
Ala Ala Val Val Pro Pro Leu Val Leu Ala Leu Ala Lys Asn Pro Leu
305                 310                 315                 320
Glu Ala Asn Phe Asp Leu Ser Ser Ile Arg Val Val Leu Ser Gly Ala
                325                 330                 335
Ala Pro Leu Gly Lys Glu Leu Glu Asp Ala Leu Arg Ser Arg Val Pro
            340                 345                 350
Gln Ala Ile Leu Gly Gln Gly Tyr Gly Met Thr Glu Ala Gly Pro Val
        355                 360                 365
Leu Ser Met Cys Leu Ala Phe Ser Lys Gln Pro Phe Pro Thr Lys Ser
    370                 375                 380
Gly Ser Cys Gly Thr Val Val Arg Asn Ala Glu Leu Lys Val Ile Asp
385                 390                 395                 400
Pro Glu Thr Gly Arg Ser Leu Gly Tyr Asn Gln Pro Gly Glu Ile Cys
                405                 410                 415
Ile Arg Gly Ser Gln Ile Met Lys Gly Tyr Leu Asn Asp Ala Glu Ala
            420                 425                 430
Thr Ala Asn Thr Ile Asp Val Glu Gly Trp Leu His Thr Gly Asp Ile
        435                 440                 445
Gly Tyr Val Asp Asp Asp Asp Glu Ile Phe Ile Val Asp Arg Val Lys
    450                 455                 460
Glu Ile Ile Lys Phe Lys Gly Phe Gln Val Pro Pro Ala Glu Leu Glu
465                 470                 475                 480
Ala Leu Leu Val Asn His Pro Ser Ile Ala Asp Ala Ala Val Val Pro
                485                 490                 495
Gln Lys Asp Glu Val Ala Gly Glu Val Pro Val Ala Phe Val Val Arg
            500                 505                 510
Ser Asp Asp Leu Asp Leu Ser Glu Glu Ala Val Lys Glu Tyr Ile Ala
        515                 520                 525
Lys Gln Val Val Phe Tyr Lys Lys Leu His Lys Val Phe Phe Val His
    530                 535                 540
Ser Ile Pro Lys Ser Ala Ser Gly Lys Ile Leu Arg Lys Asp Leu Arg
545                 550                 555                 560
Ala Lys Leu Ala Thr Ala Thr Thr Met Ser
                565                 570
```

<210> 5
<211> 1172
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 5
tgtaggattg gtggaatggg atcattccta atcccttaat gacggtggca tgaacacaaa 60
gcaaagagaa gttaggtcac tcctccttta tatatatata tatatgcatg catgaggacc 120
atggctatga tgaaggttaa tagaggtagt tgtgattgag atatgtccag cactagtttt 180
ttgttggtgt gatttctcat gatgacgcga aaatttata tatatatata atgaataata 240
tgattgatta ttctctgtaa ttttgtgaaa tagattaaaa cagctcaatg tgaggtgacc 300
agttgtcaaa tgaccactcg acttggggca tggtgatttt tcaaatcaca actcaatttg 360
aaaactaaaa ttaaaaaaga tttagattat taaattatta ggttaattca cgggttggct 420
aatcaattat tattaattaa aacgatagta tttttgataa tttaattaaa attttattgg 480
atttgaatga actcaattac atcacaaaaa acctaatcaa attaatatct tatgtgatat 540
aatttagaaa tataaatgat taacctttaa atctcgagtt tctcttataa aaaacacgta 600
taattgggct agatttaaca gctattattc aaactggcca ggacaattat taaaattaat 660
aattattatt ttttctaata aagcacttcc taattgttaa aatatatgtc taaacactaa 720
taataaaatt tatttgtgta tctttggcag taggtgagag gtgctgacaa ataaattagt 780
gcataaaata taatggattg gtggtctgtg aaaagacagg tggaggacaa gccacctctc 840
tcaagtcaaa aggccatttc acaaccaacc caaatgggaa cccaccaccg ttccccgcca 900
ttaaaatccc taatctcacc aacccaactc cacagattct tcaccaaacg caactgattt 960
ttcaatcaat gttttcccta tactaccccc ccaacaactc cataataccc aatttgtcct 1020
ttcaccaacc cccgtcctcc gtgccagcca attctatatc agcaggaatg ctctgcactc 1080
tgctttctca ggtctcctac cataagaaaa cagagagcac ctaaaactcg ccatctctcc 1140
ctctgcatct ttagcccgca atggacgcga ca 1172

<210> 6
<211> 1180
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 6
aagctttgag tattcatatg ggtattcatc cgaccattat ttttcaattt gtgttgtgtt 60
gatccaattt tcaacttatt ttttttcac ttatttttta ttagttattt ttatttttat 120
tatttttta aaaatttaaa aattaaatta taacattttt attttatccc tcattaacta 180
aaatagggat ggtaatagat attcatgaag ggagttatat atcaaatgat attagttaag 240
ctattttgat atttataccc tactcattac ttatggaata aaaaatttag atatttataa 300
aatatttatc ggatttcagg tattcatatg aatatttatt tgattattat ttattcaaca 360
aaaaataaaa caattaatat gcatgtttga agtttatata tatattaagt taggtttaga 420
tagattttgg gtggggttaa ttaatattca taccctatct actatctatc aaataatcca 480
aataaattca cctaaattag gttgggtttg tattcatcaa gttaacatta aattgtaatt 540
ccgtaagtaa ctaaacaagt acaaagactt ctattttatc ttatatatta ccataaagcc 600
aactatattt cctattcttt ttcatccctt ctatcgtaat tttctgtgac ttttttattt 660
atatattaac ggtaacgaaa cacagcaata aaagttattg tgaaagatat ggataattat 720
tatggtgact atgaaagagt aaatttgcca tgcactaagt tcctagtgtc atctcataaa 780
agacttgtct gccacgtaag ctgttgtgag tgtcgtttat ttacgcgtgt caaccaatcg 840
ctgccaattg actcttgagg gtaggtgaga gcttcggctt tgatgggaac tgcatgaggc 900
atagggtttg gtttcttgaa tgtgagatgg gcatgctttg gctcccttgc tactcacctc 960
atcttcaatt tgccagctca gctaccagtc tctcaccact agtttcacca aactttctct 1020
gctcctgtat ttattacacc ttgctcgatt ggctccgtcc tcgtacacgc atccacaccg 1080
atcgatcgat tagaaccata cagaattggg attggttggg tttacattct gcgttagata 1140
catctatcac agaaagaaac tcccttccat ctcaggaaac 1180

<210> 7
<211> 12
<212> PRT
<213> Populus tremuloides Michx. (aspen)

<400> 7
Leu Pro Tyr Ser Ser Gly Thr Thr Gly Leu Pro Lys
1 5 10

<210> 8
<211> 7

<212> PRT
<213> Populus tremuloides Michx. (aspen)

<400> 8
Gly Glu Ile Cys Ile Arg Gly
1 5

<210> 9
<211> 31
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 9
ttggatccgg iaciaciggi yticciaarg g 31

<210> 10
<211> 28
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 10
ttggatccgt igcicarcar gtigaygg 28

<210> 11
<211> 27
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 11
atgtcgacci ckdatrcada tytcicc 25

<210> 12
<211> 7
<212> PRT
<213> Populus tremuloides Michx. (aspen) unknown

<400> 12
Gly Glu Ile Cys Ile Arg Gly
1 5

<210> 13
<211> 27
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 13
tctgtctaga tgatgtcgtg gccacgg 27

<210> 14
<211> 26
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 14
ttagatctct aggacatggt ggtggc 26

<210> 15
<211> 16
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 15
cctttcacca accccc 16

<210> 16
<211> 6
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 16
ccgttc 6

<210> 17
<211> 11
<212> DNA
<213> Populus tremuloides Michx. (aspen)

<400> 17
tctcaccaac c 11

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 9/02, 9/04, 9/00, 9/24, 9/10, 9/88, 9/08, 15/82

A2

(11) International Publication Number: WO 98/11205

(43) International Publication Date: 19 March 1998 (19.03.98)

(21) International Application Number: PCT/NZ97/00112

(22) International Filing Date: 10 September 1997 (10.09.97)

(30) Priority Data:
08/713,000 11 September 1996 (11.09.96) US

(71) Applicants: GENESIS RESEARCH & DEVELOPMENT CORPORATION LIMITED [NZ/NZ]; 1 Fox Street, Parnell, Auckland (NZ). FLETCHER CHALLENGE FORESTS LIMITED [NZ/NZ]; 585 Great South Road, Penrose, Auckland (NZ).

(72) Inventors: BLOKSBERG, Leonard, Nathan; 5A Korau Road, Greenlane, Auckland (NZ). GRIERSON, Alistair, Wallace; 1/24 Medina Place, Bucklands Beach, Auckland (NZ). HAVUKKALA, Ilkka, Jaakko; 3/121 Atkin Avenue, Mission Bay, Auckland (NZ).

(74) Agents: BENNETT, Michael, Roy et al.; Russell McVeagh West-Walker, The Todd Building, Level 5, 171-177 Lambton Quay, Wellington 6001 (NZ).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: MATERIALS AND METHODS FOR THE MODIFICATION OF PLANT LIGNIN CONTENT

(57) Abstract

Novel isolated DNA sequences associated with the lignin biosynthetic pathway are provided, together with DNA constructs including such sequences. Methods for the modulation of lignin content in plants are also disclosed, the methods comprising incorporating one or more of the inventive DNA sequences or a sequence complementary to an inventive DNA sequence into the genome of a plant.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

MATERIALS AND METHODS FOR THE MODIFICATION OF PLANT LIGNIN CONTENT

Technical Field of the Invention

This invention relates to the field of modification of lignin content and composition in plants. More particularly, this invention relates to enzymes involved in the lignin biosynthetic pathway and nucleotide sequences encoding such enzymes.

Background of the Invention

Lignin is an insoluble polymer which is primarily responsible for the rigidity of plant stems. Specifically, lignin serves as a matrix around the polysaccharide components of some plant cell walls. The higher the lignin content, the more rigid the plant. For example, tree species synthesize large quantities of lignin, with lignin constituting between 20% to 30% of the dry weight of wood. In addition to providing rigidity, lignin aids in water transport within plants by rendering cell walls hydrophobic and water impermeable. Lignin also plays a role in disease resistance of plants by impeding the penetration and propagation of pathogenic agents.

The high concentration of lignin in trees presents a significant problem in the paper industry wherein considerable resources must be employed to separate lignin from the cellulose fiber needed for the production of paper. Methods typically employed for the removal of lignin are highly energy- and chemical-intensive, resulting in increased costs and increased levels of undesirable waste products. In the U.S. alone, about 20 million tons of lignin are removed from wood per year.

Lignin is largely responsible for the digestibility, or lack thereof, of forage crops, with small increases in plant lignin content resulting in relatively high decreases in digestibility. For example, crops with reduced lignin content provide more efficient forage for cattle, with the yield of milk and meat being higher relative to the amount of forage crop consumed. During normal plant growth, the increase in dry matter content is accompanied by a corresponding decrease in digestibility. When deciding on the optimum time to harvest forage crops, farmers must therefore chose between a high yield of less digestible material and a lower yield of more digestible material.

For some applications, an increase in lignin content is desirable since increasing the lignin content of a plant would lead to increased mechanical strength of wood, changes in its color and increased resistance to rot. Mycorrhizal species composition and abundance may also be favorably manipulated by modifying lignin content and structural composition.

As discussed in detail below, lignin is formed by polymerization of at least three different monolignols which are synthesized in a multistep pathway, each step in the pathway being catalyzed by a different enzyme. It has been shown that manipulation of the number of copies of genes encoding certain enzymes, such as cinnamyl alcohol dehydrogenase (CAD) and caffeic acid 3-O-methyltransferase (COMT) results in modification of the amount of lignin produced; see, for example, U.S. Patent No. 5,451,514 and PCT publication no. WO 94/23044. Furthermore, it has been shown that antisense expression of sequences encoding CAD in poplar leads to the production of lignin having a modified composition (Grand, C. et al. Planta (Berl.) 163:232-237 (1985)).

While DNA sequences encoding some of the enzymes involved in the lignin biosynthetic pathway have been isolated for certain species of plants, genes encoding many of the enzymes in a wide range of plant species have not yet been identified. Thus there remains a need in the art for materials useful in the modification of lignin content and composition in plants and for methods for their use.

Summary of the Invention

Briefly, the present invention provides isolated DNA sequences obtainable from eucalyptus and pine which encode enzymes involved in the lignin biosynthetic pathway, DNA constructs including such sequences, and methods for the use of such constructs. Transgenic plants having altered lignin content and composition are also provided.

In a first aspect, the present invention provides isolated DNA sequences coding for the following enzymes isolated from eucalyptus and pine: cinnamate 4-hydroxylase (C4H), coumarate 3-hydroxylase (C3H), phenolase (PNL), O-methyl transferase (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL), coniferol

glucosyl transferase (CGT), coniferin *beta*-glucosidase (CBG), laccase (LAC) and peroxidase (POX), together with ferulate-5-hydroxylase (F5H) from eucalyptus. In one embodiment, the isolated DNA sequences comprise a nucleotide sequence selected from the group consisting of: (a) sequences recited in SEQ ID NO: 3, 13, 16-70, and 72-88; (b) complements of the sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88; (c) reverse complements of the sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88; (d) reverse sequences of the sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88; and (e) sequences having at least about a 99% probability of being the same as a sequence of (a) – (d) as measured by the computer algorithm FASTA.

In another aspect, the invention provides DNA constructs comprising a DNA sequence of the present invention, either alone, in combination with one or more of the inventive sequences or in combination with one or more known DNA sequences; together with transgenic cells comprising such constructs.

In a related aspect, the present invention provides DNA constructs comprising, in the 5'-3' direction, a gene promoter sequence; an open reading frame coding for at least a functional portion of an enzyme encoded by the inventive DNA sequences or variants thereof; and a gene termination sequence. The open reading frame may be orientated in either a sense or antisense direction. DNA constructs comprising a non-coding region of a gene coding for an enzyme encoded by the above DNA sequences or a nucleotide sequence complementary to a non-coding region, together with a gene promoter sequence and a gene termination sequence, are also provided. Preferably, the gene promoter and termination sequences are functional in a host plant. Most preferably, the gene promoter and termination sequences are those of the original enzyme genes but others generally used in the art, such as the Cauliflower Mosaic Virus (CMV) promoter, with or without enhancers, such as the Kozak sequence or Omega enhancer, and *Agrobacterium tumefaciens* nopalin synthase terminator may be usefully employed in the present invention. Tissue-specific promoters may be employed in order to target expression to one or more desired tissues. In a preferred embodiment, the gene promoter sequence provides for transcription in xylem. The DNA construct may further include a marker for the identification of transformed cells.

In a further aspect, transgenic plant cells comprising the DNA constructs of the present invention are provided, together with plants comprising such transgenic cells, and fruits and seeds of such plants.

In yet another aspect, methods for modulating the lignin content and composition of a plant are provided, such methods including stably incorporating into the genome of the plant a DNA construct of the present invention. In a preferred embodiment, the target plant is a woody plant, preferably selected from the group consisting of eucalyptus and pine species, most preferably from the group consisting of *Eucalyptus grandis* and *Pinus radiata*. In a related aspect, a method for producing a plant having altered lignin content is provided, the method comprising transforming a plant cell with a DNA construct of the present invention to provide a transgenic cell, and cultivating the transgenic cell under conditions conducive to regeneration and mature plant growth.

In yet a further aspect, the present invention provides methods for modifying the activity of an enzyme in a plant, comprising stably incorporating into the genome of the plant a DNA construct of the present invention. In a preferred embodiment, the target plant is a woody plant, preferably selected from the group consisting of eucalyptus and pine species, most preferably from the group consisting of *Eucalyptus grandis* and *Pinus radiata*.

The above-mentioned and additional features of the present invention and the manner of obtaining them will become apparent, and the invention will be best understood by reference to the following more detailed description, read in conjunction with the accompanying drawing.

Brief Description of the Figures

Fig. 1 is a schematic overview of the lignin biosynthetic pathway.

Detailed Description

Lignin is formed by polymerization of at least three different monolignols, primarily *para*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol. While these three types of lignin subunits are well known, it is possible that slightly different variants of these subunits may be involved in the lignin biosynthetic pathway in various

plants. The relative concentration of these residues in lignin varies between different plant species and within species. In addition, the composition of lignin may also vary between different tissues within a specific plant. The three monolignols are derived from phenylalanine in a multistep process and are believed to be polymerized into lignin by a free radical mechanism.

Fig. 1 shows the different steps in the biosynthetic pathway for coniferyl alcohol together with the enzymes responsible for catalyzing each step. *para*-Coumaryl alcohol and sinapyl alcohol are synthesized by similar pathways. Phenylalanine is first deaminated by phenylalanine ammonia-lyase (PAL) to give cinnamate which is then hydroxylated by cinnamate 4-hydroxylase (C4H) to form *p*-coumarate. *p*-Coumarate is hydroxylated by coumarate 3-hydroxylase to give caffeate. The newly added hydroxyl group is then methylated by O-methyl transferase (OMT) to give ferulate which is conjugated to coenzyme A by 4-coumarate:CoA ligase (4CL) to form feruloyl-CoA. Reduction of feruloyl-CoA to coniferaldehyde is catalyzed by cinnamoyl-CoA reductase (CCR). Coniferaldehyde is further reduced by the action of cinnamyl alcohol dehydrogenase (CAD) to give coniferyl alcohol which is then converted into its glucosylated form for export from the cytoplasm to the cell wall by coniferol glucosyl transferase (CGT). Following export, the de-glucosylated form of coniferyl alcohol is obtained by the action of coniferin *beta*-glucosidase (CBG). Finally, polymerization of the three monolignols to provide lignin is catalyzed by phenolase (PNL), laccase (LAC) and peroxidase (POX).

The formation of sinapyl alcohol involves an additional enzyme, ferulate-5-hydroxylase (F5H). For a more detailed review of the lignin biosynthetic pathway, see: Whetton, R. and Sederoff, R., The Plant Cell, 7:1001-1013 (1995).

Quantitative and qualitative modifications in plant lignin content are known to be induced by external factors such as light stimulation, low calcium levels and mechanical stress. Synthesis of new types of lignins, sometimes in tissues not normally lignified, can also be induced by infection with pathogens. In addition to lignin, several other classes of plant products are derived from phenylalanine, including flavonoids, coumarins, stilbenes and benzoic acid derivatives, with the initial steps in the synthesis of all these compounds being the same. Thus modification of the action of PAL, C4H and 4CL may affect the synthesis of other plant products in addition to lignin.

Using the methods and materials of the present invention, the lignin content of a plant can be increased by incorporating additional copies of genes encoding enzymes involved in the lignin biosynthetic pathway into the genome of the target plant. Similarly, a decrease in lignin content can be obtained by transforming the target plant with antisense copies of such genes. In addition, the number of copies of genes encoding for different enzymes in the lignin biosynthetic pathway can be manipulated to modify the relative amount of each monolignol synthesized, thereby leading to the formation of lignin having altered composition. The alteration of lignin composition would be advantageous, for example, in tree processing for paper, and may also be effective in altering the palatability of wood materials to rotting fungi.

In one embodiment, the present invention provides isolated complete or partial DNA sequences encoding, or partially encoding, enzymes involved in the lignin biosynthetic pathway, the DNA sequences being obtainable from eucalyptus and pine. Specifically, the present invention provides isolated DNA sequences encoding the enzymes CAD (SEQ ID NO: 1, 30), PAL (SEQ ID NO: 16), C4H (SEQ ID NO: 17), C3H (SEQ ID NO: 18), F5H (SEQ ID NO: 19-21), OMT (SEQ ID NO: 22-25), CCR (SEQ ID NO: 26-29), CGT (SEQ ID NO: 31-33), CBG (SEQ ID NO: 34), PNL (SEQ ID NO: 35, 36), LAC (SEQ ID NO: 37-41) and POX (SEQ ID NO: 42-44) from *Eucalyptus grandis*; and the enzymes C4H (SEQ ID NO: 2, 3, 48, 49), C3H (SEQ ID NO: 4, 50-52), PNL (SEQ ID NO: 5, 81), OMT (SEQ ID NO: 6, 53-55), CAD (SEQ ID NO: 7, 71), CCR (SEQ ID NO: 8, 58-70), PAL (SEQ ID NO: 9-11,45-47), 4CL (SEQ ID NO: 12, 56, 57), CGT (SEQ ID NO: 72), CBG (SEQ ID NO: 73-80), LAC (SEQ ID NO: 82-84) and POX (SEQ ID NO: 13, 85-88) from *Pinus radiata*. Complements of such isolated DNA sequences, reverse complements of such isolated DNA sequences and reverse sequences of such isolated DNA sequences, together with variants of such sequences, are also provided. DNA sequences encompassed by the present invention include cDNA, genomic DNA, recombinant DNA and wholly or partially chemically synthesized DNA molecules.

The definition of the terms "complement", "reverse complement" and "reverse sequence", as used herein, is best illustrated by the following example. For the sequence 5' AGGACC 3', the complement, reverse complement and reverse sequence are as follows:

complement	3' TCCTGG 5'
reverse complement	3' GGTCCT 5'
reverse sequence	5' CCAGGA 3'.

As used herein, the term "variant" covers any sequence which exhibits at least about 50%, more preferably at least about 70% and, more preferably yet, at least about 90% identity to a sequence of the present invention. Most preferably, a "variant" is any sequence which has at least about a 99% probability of being the same as the inventive sequence. The probability for DNA sequences is measured by the computer algorithm FASTA (version 2.0u4, February 1996; Pearson W. R. et al., Proc. Natl. Acad. Sci., 85:2444-2448, 1988), the probability for translated DNA sequences is measured by the computer algorithm TBLASTX and that for protein sequences is measured by the computer algorithm BLASTP (Altschul, S. F. et al. J. Mol. Biol., 215:403-410, 1990). The term "variants" thus encompasses sequences wherein the probability of finding a match by chance (smallest sum probability) in a database, is less than about 1% as measured by any of the above tests.

Variants of the isolated sequences from other eucalyptus and pine species, as well as from other commercially important species utilized by the lumber industry, are contemplated. These include the following gymnosperms, by way of example: loblolly pine *Pinus taeda*, slash pine *Pinus elliotti*, sand pine *Pinus clausa*, longleaf pine *Pinus palustrus*, shortleaf pine *Pinus echinata*, ponderosa pine *Pinus ponderosa*, Jeffrey pine *Pinus jeffrey*, red pine *Pinus resinosa*, pitch pine *Pinus rigida*, jack pine *Pinus banksiana*, pond pine *Pinus serotina*, Eastern white pine *Pinus strobus*, Western white pine *Pinus monticola*, sugar pine *Pinus lambertiana*, Virginia pine *Pinus virginiana*, lodgepole pine *Pinus contorta*, Caribbean pine *Pinus caribaea*, *P. pinaster*, Calabrian pine *P. brutia*, Afghan pine *P. eldarica*, Coulter pine *P. coulteri*, European pine *P. nigra* and *P. sylvestris*; Douglas-fir *Pseudotsuga menziesii*; the hemlocks which include Western hemlock *Tsuga heterophylla*, Eastern hemlock *Tsuga canadensis*, Mountain hemlock *Tsuga mertensiana*; the spruces which include the Norway spruce *Picea abies*, red spruce *Picea rubens*, white spruce *Picea glauca*, black spruce *Picea mariana*, Sitka spruce *Picea sitchensis*, Englemann spruce *Picea engelmanni*, and blue spruce *Picea pungens*; redwood *Sequoia sempervirens*; the true firs include the Alpine fir *Abies lasiocarpa*, silver fir *Abies amabilis*, grand fir *Abies grandis*, noble fir *Abies procera*, white fir *Abies concolor*, California red fir *Abies magnifica*, and balsam fir *Abies balsamea*, the cedars which include the Western red cedar *Thuja plicata*, incense

cedar *libocedrus decurrens*, Northern white cedar *Thuja occidentalis*, Port Orford cedar *Chamaecyparis lawsoniona*, Atlantic white cedar *Chamaecyparis thyoides*, Alaska yellow-cedar *Chamaecyparis nootkatensis*, and Eastern red cedar *Huniperus virginiana*; the larches which include Eastern larch *Larix laricina*, Western larch *Larix occidentalis*, European larch *Larix decidua*, Japanese larch *Larix leptolepis*, and Siberian larch *Larix siberica*; bold cypress *Taxodium distichum* and Giant sequoia *Sequoia gigantea*;

and the following angiosperms, by way of example:

Eucalyptus alba, E. bancroftii, E. botyroides, E. bridgesiana, E. calophylla, E. camaldulensis, E. citriodora, E. cladocalyx, E. coccifera, E. curtisii, E. dalrympleana, E. deglupta, E. delagatensis, E. diversicolor, E. dunnii, E. ficifolia, E. globulus, E. gomphocephala, E. gunnii, E. henryi, E. laevopinea, E. macarthurii, E. macrorhyncha, E. maculata, E. marginata, E. megacarpa, E. melliodora, E. nicholii, E. nitens, E. nova-anglica, E. obliqua, E. obtusiflora, E. oreades, E. pauciflora, E. polybractea, E. regnans, E. resinifera, E. robusta, E. rudis, E. saligna, E. sideroxylon, E. stuartiana, E. tereticornis, E. torelliana, E. urnigera, E. urophylla, E. viminalis, E. viridis, E. wandoo and *E. youmanni*.

The inventive DNA sequences may be isolated by high throughput sequencing of cDNA libraries such as those prepared from *Eucalyptus grandis* and *Pinus radiata* as described below in Examples 1 and 2. Alternatively, oligonucleotide probes based on the sequences provided in SEQ ID NO: 1-13 and 16-88 can be synthesized and used to identify positive clones in either cDNA or genomic DNA libraries from *Eucalyptus grandis* and *Pinus radiata*, or from other gymnosperms and angiosperms including those identified above, by means of hybridization or PCR techniques. Probes can be shorter than the sequences provided herein but should be at least about 10, preferably at least about 15 and most preferably at least about 20 nucleotides in length. Hybridization and PCR techniques suitable for use with such oligonucleotide probes are well known in the art. Positive clones may be analyzed by restriction enzyme digestion, DNA sequencing or the like.

In addition, the DNA sequences of the present invention may be generated by synthetic means using techniques well known in the art. Equipment for automated synthesis of oligonucleotides is commercially available from suppliers such as Perkin Elmer/Applied Biosystems Division (Foster City, CA) and may be operated according to the manufacturer's instructions.

In one embodiment, the DNA constructs of the present invention include an open reading frame coding for at least a functional portion of an enzyme encoded by a nucleotide sequence of the present invention or a variant thereof. As used herein, the "functional portion" of an enzyme is that portion which contains the active site essential for affecting the metabolic step, *i.e.* the portion of the molecule that is capable of binding one or more reactants or is capable of improving or regulating the rate of reaction. The active site may be made up of separate portions present on one or more polypeptide chains and will generally exhibit high substrate specificity. The term "enzyme encoded by a nucleotide sequence" as used herein, includes enzymes encoded by a nucleotide sequence which includes the partial isolated DNA sequences of the present invention.

For applications where amplification of lignin synthesis is desired, the open reading frame is inserted in the DNA construct in a sense orientation, such that transformation of a target plant with the DNA construct will lead to an increase in the number of copies of the gene and therefore an increase in the amount of enzyme. When down-regulation of lignin synthesis is desired, the open reading frame is inserted in the DNA construct in an antisense orientation, such that the RNA produced by transcription of the DNA sequence is complementary to the endogenous mRNA sequence. This, in turn, will result in a decrease in the number of copies of the gene and therefore a decrease in the amount of enzyme. Alternatively, regulation can be achieved by inserting appropriate sequences or subsequences (e.g. DNA or RNA) in ribozyme constructs.

In a second embodiment, the inventive DNA constructs comprise a nucleotide sequence including a non-coding region of a gene coding for an enzyme encoded by a DNA sequence of the present invention, or a nucleotide sequence complementary to such a non-coding region. As used herein the term "non-coding region" includes both transcribed sequences which are not translated, and non-transcribed sequences within about 2000 base pairs 5' or 3' of the translated sequences or open reading frames. Examples of non-coding regions which may be usefully employed in the inventive constructs include introns and 5'-non-coding leader sequences. Transformation of a target plant with such a DNA construct may lead to a reduction in the amount of lignin synthesized by the plant by the process of cosuppression, in a manner similar to that

discussed, for example, by Napoli et al. (Plant Cell 2:279-290, 1990) and de Carvalho Niebel et al. (Plant Cell 7:347-358, 1995).

The DNA constructs of the present invention further comprise a gene promoter sequence and a gene termination sequence, operably linked to the DNA sequence to be transcribed, which control expression of the gene. The gene promoter sequence is generally positioned at the 5' end of the DNA sequence to be transcribed, and is employed to initiate transcription of the DNA sequence. Gene promoter sequences are generally found in the 5' non-coding region of a gene but they may exist in introns (Luehrsen, K. R., Mol. Gen. Genet. 225:81-93, 1991) or in the coding region, as for example in PAL of tomato (Bloksberg, 1991, Studies on the Biology of Phenylalanine Ammonia Lyase and Plant Pathogen Interaction. Ph.D. Thesis, Univ. of California, Davis, University Microfilms International order number 9217564). When the construct includes an open reading frame in a sense orientation, the gene promoter sequence also initiates translation of the open reading frame. For DNA constructs comprising either an open reading frame in an antisense orientation or a non-coding region, the gene promoter sequence consists only of a transcription initiation site having a RNA polymerase binding site.

A variety of gene promoter sequences which may be usefully employed in the DNA constructs of the present invention are well known in the art. The promoter gene sequence, and also the gene termination sequence, may be endogenous to the target plant host or may be exogenous, provided the promoter is functional in the target host. For example, the promoter and termination sequences may be from other plant species, plant viruses, bacterial plasmids and the like. Preferably, gene promoter and termination sequences are from the inventive sequences themselves.

Factors influencing the choice of promoter include the desired tissue specificity of the construct, and the timing of transcription and translation. For example, constitutive promoters, such as the 35S Cauliflower Mosaic Virus (CaMV 35S) promoter, will affect the activity of the enzyme in all parts of the plant. Use of a tissue specific promoter will result in production of the desired sense or antisense RNA only in the tissue of interest. With DNA constructs employing inducible gene promoter sequences, the rate of RNA polymerase binding and initiation can be modulated by external stimuli, such as light, heat, anaerobic stress, alteration in nutrient conditions

and the like. Temporally regulated promoters can be employed to effect modulation of the rate of RNA polymerase binding and initiation at a specific time during development of a transformed cell. Preferably, the original promoters from the enzyme gene in question, or promoters from a specific tissue-targeted gene in the organism to be transformed, such as eucalyptus or pine are used. Other examples of gene promoters which may be usefully employed in the present invention include, mannopine synthase (mas), octopine synthase (ocs) and those reviewed by Chua et al. (Science, 244:174-181, 1989).

The gene termination sequence, which is located 3' to the DNA sequence to be transcribed, may come from the same gene as the gene promoter sequence or may be from a different gene. Many gene termination sequences known in the art may be usefully employed in the present invention, such as the 3' end of the *Agrobacterium tumefaciens* nopaline synthase gene. However, preferred gene terminator sequences are those from the original enzyme gene or from the target species to be transformed.

The DNA constructs of the present invention may also contain a selection marker that is effective in plant cells, to allow for the detection of transformed cells containing the inventive construct. Such markers, which are well known in the art, typically confer resistance to one or more toxins. One example of such a marker is the NPTII gene whose expression results in resistance to kanamycin or hygromycin, antibiotics which is usually toxic to plant cells at a moderate concentration (Rogers et al. in Methods for Plant Molecular Biology, A. Weissbach and H. Weissbach, eds., Academic Press Inc., San Diego, CA (1988)). Alternatively, the presence of the desired construct in transformed cells can be determined by means of other techniques well known in the art, such as Southern and Western blots.

Techniques for operatively linking the components of the inventive DNA constructs are well known in the art and include the use of synthetic linkers containing one or more restriction endonuclease sites as described, for example, by Maniatis et al., (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). The DNA construct of the present invention may be linked to a vector having at least one replication system, for example, *E. coli*, whereby after each manipulation, the resulting construct can be cloned and sequenced and the correctness of the manipulation determined.

The DNA constructs of the present invention may be used to transform a variety of plants, both monocotyledonous (e.g. grasses, corn, grains, oat, wheat and barley), dicotyledonous (e.g. *Arabidopsis*, tobacco, legumes, alfalfa, oaks, eucalyptus, maple), and Gymnosperms (e.g. Scots pine (Aronen, Finnish Forest Res. Papers, vol. 595, 1996), white spruce (Ellis et al., Biotechnology 11:94-92, 1993), larch (Huang et al., In Vitro Cell 27:201-207, 1991). In a preferred embodiment, the inventive DNA constructs are employed to transform woody plants, herein defined as a tree or shrub whose stem lives for a number of years and increases in diameter each year by the addition of woody tissue. Preferably the target plant is selected from the group consisting of eucalyptus and pine species, most preferably from the group consisting of *Eucalyptus grandis* and *Pinus radiata*. As discussed above, transformation of a plant with a DNA construct including an open reading frame coding for an enzyme encoded by an inventive DNA sequence wherein the open reading frame is orientated in a sense direction will lead to an increase in lignin content of the plant or, in some cases, to a decrease by cosuppression. Transformation of a plant with a DNA construct comprising an open reading frame in an antisense orientation or a non-coding (untranslated) region of a gene will lead to a decrease in the lignin content of the transformed plant.

Techniques for stably incorporating DNA constructs into the genome of target plants are well known in the art and include *Agrobacterium tumefaciens* mediated introduction, electroporation, protoplast fusion, injection into reproductive organs, injection into immature embryos, high velocity projectile introduction and the like. The choice of technique will depend upon the target plant to be transformed. For example, dicotyledonous plants and certain monocots and gymnosperms may be transformed by *Agrobacterium* Ti plasmid technology, as described, for example by Bevan (Nucl. Acid Res. 12:8711-8721, 1984). Targets for the introduction of the DNA constructs of the present invention include tissues, such as leaf tissue, disseminated cells, protoplasts, seeds, embryos, meristematic regions; cotyledons, hypocotyls, and the like. One preferred method for transforming eucalyptus and pine is a biolistic method using pollen (see, for example, Aronen 1996, Finnish Forest Res. Papers vol. 595, 53pp) or easily regenerable embryonic tissues. Other transformation techniques which may be usefully employed in the inventive methods include those taught by Ellis et al. (Plant

Cell Reports, 8:16-20, 1989), Wilson et al. (Plant Cell Reports 7:704-707, 1989) and Tautorus et al. (Theor. Appl. Genet. 78:531-536, 1989).

Once the cells are transformed, cells having the inventive DNA construct incorporated in their genome may be selected by means of a marker, such as the kanamycin resistance marker discussed above. Transgenic cells may then be cultured in an appropriate medium to regenerate whole plants, using techniques well known in the art. In the case of protoplasts, the cell wall is allowed to reform under appropriate osmotic conditions. In the case of seeds or embryos, an appropriate germination or callus initiation medium is employed. For explants, an appropriate regeneration medium is used. Regeneration of plants is well established for many species. For a review of regeneration of forest trees see Dunstan et al., Somatic embryogenesis in woody plants. In: Thorpe, T.A. ed., 1995: in vitro embryogenesis of plants. Vol. 20 in Current Plant Science and Biotechnology in Agriculture, Chapter 12, pp. 471-540. Specific protocols for the regeneration of spruce are discussed by Roberts et al., (Somatic Embryogenesis of Spruce. In: *Synseed. Applications of synthetic seed to crop improvement.* Redenbaugh, K., ed. CRC Press, Chapter 23, pp. 427-449, 1993). The resulting transformed plants may be reproduced sexually or asexually, using methods well known in the art, to give successive generations of transgenic plants.

As discussed above, the production of RNA in target plant cells can be controlled by choice of the promoter sequence, or by selecting the number of functional copies or the site of integration of the DNA sequences incorporated into the genome of the target plant host. A target plant may be transformed with more than one DNA construct of the present invention, thereby modulating the lignin biosynthetic pathway for the activity of more than one enzyme, affecting enzyme activity in more than one tissue or affecting enzyme activity at more than one expression time. Similarly, a DNA construct may be assembled containing more than one open reading frame coding for an enzyme encoded by a DNA sequence of the present invention or more than one non-coding region of a gene coding for such an enzyme. The DNA sequences of the present inventive may also be employed in combination with other known sequences encoding enzymes involved in the lignin biosynthetic pathway. In this manner, it may be possible to add a lignin biosynthetic pathway to a non-woody plant to produce a new woody plant.

The isolated DNA sequences of the present invention may also be employed as probes to isolate DNA sequences encoding enzymes involved in the lignin synthetic pathway from other plant species, using techniques well known to those of skill in the art.

The following examples are offered by way of illustration and not by way of limitation.

Example 1

Isolation and Characterization of cDNA Clones from *Eucalyptus grandis*

Two *Eucalyptus grandis* cDNA expression libraries (one from a mixture of various tissues from a single tree and one from leaves of a single tree) were constructed and screened as follows.

mRNA was extracted from the plant tissue using the protocol of Chang et al. (Plant Molecular Biology Reporter 11:113-116 (1993)) with minor modifications. Specifically, samples were dissolved in CPC-RNAXB (100 mM Tris-Cl, pH 8,0; 25 mM EDTA; 2.0 M NaCl; 2%CTAB; 2% PVP and 0.05% Spermidine*3 HCl)and extracted with Chloroform:isoamyl alcohol, 24:1. mRNA was precipitated with ethanol and the total RNA preparate was purified using a Poly(A) Quik mRNA Isolation Kit (Stratagene, La Jolla, CA). A cDNA expression library was constructed from the purified mRNA by reverse transcriptase synthesis followed by insertion of the resulting cDNA clones in Lambda ZAP using a ZAP Express cDNA Synthesis Kit (Stratagene), according to the manufacturer's protocol. The resulting cDNAs were packaged using a Gigapack II Packaging Extract (Stratagene) employing 1 µl of sample DNA from the 5 µl ligation mix. Mass excision of the library was done using XL1-Blue MRF' cells and XLOLR cells (Stratagene) with ExAssist helper phage (Stratagene). The excised phagemids were diluted with NZY broth (Gibco BRL, Gaithersburg, MD) and plated out onto LB-kanamycin agar plates containing X-gal and isopropylthio-beta-galactoside (IPTG).

Of the colonies plated and picked for DNA miniprep, 99% contained an insert suitable for sequencing. Positive colonies were cultured in NZY broth with kanamycin and cDNA was purified by means of alkaline lysis and polyethylene glycol (PEG) precipitation. Agarose gel at 1% was used to screen sequencing templates for

chromosomal contamination. Dye primer sequences were prepared using a Turbo Catalyst 800 machine (Perkin Elmer/Applied Biosystems, Foster City, CA) according to the manufacturer's protocol.

DNA sequence for positive clones was obtained using an Applied Biosystems Prism 377 sequencer. cDNA clones were sequenced first from both the 5' end and, in some cases, also from the 3' end. For some clones, internal sequence was obtained using subcloned fragments. Subcloning was performed using standard procedures of restriction mapping and subcloning to pBluescript II SK+ vector.

The determined cDNA sequence was compared to known sequences in the EMBL database (release 46, March 1996) using the FASTA algorithm of February 1996 (version 2.0u4) (available on the Internet at the ftp site ftp://ftp.virginia.edu/pub/fasta/). Multiple alignments of redundant sequences were used to build up reliable consensus sequences. Based on similarity to known sequences from other plant species, the isolated DNA sequence (SEQ ID NO: 1) was identified as encoding a CAD enzyme.

In further studies, using the procedure described above, cDNA sequences encoding the following *Eucalyptus grandis* enzymes were isolated: PAL (SEQ ID NO: 16); C4H (SEQ ID NO: 17); C3H (SEQ ID NO: 18); F5H (SEQ ID NO: 19-21); OMT (SEQ ID NO: 22-25); CCR (SEQ ID NO: 26-29); CAD (SEQ ID NO: 30); CGT (SEQ ID NO: 31-33); CBG (SEQ ID NO: 34); PNL (SEQ ID NO: 35, 36); LAC (SEQ ID NO: 37-41); and POX (SEQ ID NO: 42-44).

Example 2

Isolation and Characterization of cDNA Clones from *Pinus radiata*

a) Isolation of cDNA clones by high through-put screening

A *Pinus radiata* cDNA expression library was constructed from xylem and screened as described above in Example 1. DNA sequence for positive clones was obtained using forward and reverse primers on an Applied Biosystems Prism 377 sequencer and the determined sequences were compared to known sequences in the database as described above.

Based on similarity to known sequences from other plant species, the isolated DNA sequences were identified as encoding the enzymes C4H (SEQ ID NO: 2 and 3), C3H (SEQ ID NO: 4), PNL (SEQ ID NO: 5), OMT (SEQ ID NO: 6), CAD (SEQ ID NO: 7), CCR (SEQ ID NO: 8), PAL (SEQ ID NO: 9-11) and 4CL (SEQ ID NO: 12).

In further studies, using the procedure described above, additional cDNA clones encoding the following *Pinus radiata* enzymes were isolated: PAL (SEQ ID NO: 45-47); C4H (SEQ ID NO: 48, 49); C3H (SEQ ID NO: 50-52); OMT (SEQ ID NO: 53-55); 4CL (SEQ ID NO: 56, 57); CCR (SEQ ID NO: 58-70); CAD (SEQ ID NO: 71); CGT (SEQ ID NO: 72); CBG (SEQ ID NO: 73-80); PNL (SEQ ID NO: 81); LAC (SEQ ID NO: 82-84); and POX (SEQ ID NO: 85-88).

b) Isolation of cDNA clones by PCR

Two PCR probes, hereinafter referred to as LNB010 and LNB011 (SEQ ID NO: 14 and 15, respectively) were designed based on conserved domains in the following peroxidase sequences previously identified in other species: vanpox, hvupox6, taepox, hvupox1, osapox, ntopox2, ntopox1, lespox, pokpox, luspox, athpox, hrpox, spopox, and tvepox (Genbank accession nos. D11337, M83671, X56011, X58396, X66125, J02979, D11396, X71593, D11102, L07554, M58381, X57564, Z22920, and Z31011, respectively).

RNA was isolated from pine xylem and first strand cDNA was synthesized as described above. This cDNA was subjected to PCR using 4 μM LNB010, 4 μM LNB011, 1 x Kogen's buffer, 0.1 mg/ml BSA, 200 mM dNTP, 2 mM Mg^{2+}, and 0.1 U/μl of Taq polymerase (Gibco BRL). Conditions were 2 cycles of 2 min at 94 °C, 1 min at 55 °C and 1 min at 72 °C; 25 cycles of 1 min at 94 °C, 1 min at 55 °C, and 1 min at 72 °C; and 18 cycles of 1 min at 94 °C, 1 min at 55 °C, and 3 min at 72 °C in a Stratagene Robocycler. The gene was re-amplified in the same manner. A band of about 200 bp was purified from a TAE agarose gel using a Schleicher & Schuell Elu-Quik DNA purification kit and clones into a T-tailed pBluescript vector (Marchuk D. et al., Nucleic Acids Res. 19:1154, 1991). Based on similarity to known sequences, the isolated gene (SEQ ID NO: 13) was identified as encoding pine peroxidase (POX).

Example 3

Use of an O-methyltransferase (OMT) Gene to Modify Lignin Biosynthesis

a) Transformation of tobacco plants with a *Pinus radiata* OMT gene

Sense and anti-sense constructs containing a sequence including the coding region of OMT (SEQ ID NO: 53) from *Pinus radiata* were inserted into *Agrobacterium tumefaciens* LBA4301 (provided as a gift by Dr. C. Kado, University of California, Davis, CA) by direct transformation using published methods (see, An G, Ebert PR, Mitra A, Ha SB: Binary Vectors. In: Gelvin SB, Schilperoort RA (eds) Plant Molecular Biology Manual, Kluwer Academic Publishers, Dordrecht (1988)). The presence and integrity of the transgenic constructs were verified by restriction digestion and DNA sequencing.

Tobacco (*Nicotiana tabacum* cv. Samsun) leaf sections were transformed using the method of Horsch et al. (Science, 227:1229-1231, 1985). Five independent transformed plant lines were established for the sense construct and eight independent transformed plant lines were established for the anti-sense construct for OMT. Transformed plants containing the appropriate lignin gene construct were verified using Southern blot experiments. A "+" in the column labeled "Southern" in Table 1 below indicates that the transformed plant lines were confirmed as independent transformed lines.

b) Expression of *Pinus* OMT in transformed plants

Total RNA was isolated from each independent transformed plant line created with the OMT sense and anti-sense constructs. The RNA samples were analysed in Northern blot experiments to determine the level of expression of the transgene in each transformed line. The data shown in the column labeled "Northern" in Table 1 shows that the transformed plant lines containing the sense and anti-sense constructs for OMT all exhibited high levels of expression, relative to the background on the Northern blots. OMT expression in sense plant line number 2 was not measured because the RNA sample showed signs of degradation. There was no detectable hybridisation to RNA samples from empty vector-transformed control plants.

c) Modulation of OMT enzyme activity in transformed plants

The total activity of OMT enzyme, encoded by the *Pinus* OMT gene and by the endogenous tobacco OMT gene, in transformed tobacco plants was analysed for each transformed plant line created with the OMT sense and anti-sense constructs. Crude protein extracts were prepared from each transformed plant and assayed using the method of Zhang et al. (Plant Physiol., 113:65-74, 1997). The data contained in the column labeled "Enzyme" in Table 1 shows that the transformed plant lines containing the OMT sense construct generally had elevated OMT enzyme activity, with a maximum of 199%, whereas the transformed plant lines containing the OMT anti-sense construct generally had reduced OMT enzyme activity, with a minimum of 35%, relative to empty vector-transformed control plants. OMT enzyme activity was not estimated in sense plant line number 3.

d) Effects of *Pinus* OMT on lignin concentration in transformed plants

The concentration of lignin in the transformed tobacco plants was determined using the well-established procedure of thioglycolic acid extraction (see, Freudenberg et al. in "Constitution and Biosynthesis of Lignin", Springer-Verlag, Berlin, 1968). Briefly, whole tobacco plants, of an average age of 38 days, were frozen in liquid nitrogen and ground to a fine powder in a mortar and pestle. 100 mg of frozen powder from one empty vector-transformed control plant line, the five independent transformed plant lines containing the sense construct for OMT and the eight independent transformed plant lines containing the anti-sense construct for OMT were extracted individually with methanol, followed by 10% thioglycolic acid and finally dissolved in 1 M NaOH. The final extracts were assayed for absorbance at 280 nm. The data shown in the column labelled "TGA" in Table 1 shows that the transformed plant lines containing the sense and the anti-sense OMT gene constructs all exhibited significantly decreased levels of lignin, relative to the empty vector-transformed control plant lines.

Table 1

plant line	transgene	orientation	Southern	Northern	Enzyme	TGA
1	control	na	+	blank	100	104
1	OMT	sense	+	2.9E+6	86	55
2	OMT	sense	+	na	162	58
3	OMT	sense	+	4.1E+6	na	63
4	OMT	sense	+	2.3E+6	142	66
5	OMT	sense	+	3.6E+5	199	75
1	OMT	anti-sense	+	1.6E+4	189	66
2	OMT	anti-sense	+	5.7E+3	35	70
3	OMT	anti-sense	+	8.0E+3	105	73
4	OMT	anti-sense	+	1.4E+4	109	74
5	OMT	anti-sense	+	2.5E+4	87	78
6	OMT	anti-sense	+	2.5E+4	58	84
7	OMT	anti-sense	+	2.5E+4	97	92
8	OMT	anti-sense	+	1.1E+4	151	94

These data clearly indicate that lignin concentration, as measured by the TGA assay, can be directly manipulated by either sense or anti-sense expression of a lignin biosynthetic gene such as OMT.

Example 4

Use of a 4-Coumarate:CoA ligase (4CL) Gene to Modify Lignin Biosynthesis

a) Transformation of tobacco plants with a *Pinus radiata* 4CL gene

Sense and anti-sense constructs containing a sequence including the coding region of 4CL (SEQ ID NO: 56) from *Pinus radiata* were inserted into *Agrobacterium tumefaciens* LBA4301 by direct transformation as described above. The presence and integrity of the transgenic constructs were verified by restriction digestion and DNA sequencing.

Tobacco (*Nicotiana tabacum* cv. Samsun) leaf sections were transformed as described above. Five independent transformed plant lines were established for the sense construct and eight independent transformed plant lines were established for the anti-sense construct for 4CL. Transformed plants containing the appropriate lignin gene construct were verified using Southern blot experiments. A "+" in the column

labeled "Southern" in Table 2 indicates that the transformed plant lines listed were confirmed as independent transformed lines.

b) Expression of *Pinus* 4CL in transformed plants

Total RNA was isolated from each independent transformed plant line created with the 4CL sense and anti-sense constructs. The RNA samples were analysed in Northern blot experiments to determine the level of expression of the transgene in each transformed line. The data shown in the column labelled "Northern" in Table 2 below shows that the transformed plant lines containing the sense and anti-sense constructs for 4CL all exhibit high levels of expression, relative to the background on the Northern blots. 4CL expression in anti-sense plant line number 1 was not measured because the RNA was not available at the time of the experiment. There was no detectable hybridisation to RNA samples from empty vector-transformed control plants.

c) Modulation of 4CL enzyme activity in transformed plants

The total activity of 4CL enzyme, encoded by the *Pinus* 4CL gene and by the endogenous tobacco 4CL gene, in transformed tobacco plants was analysed for each transformed plant line created with the 4CL sense and anti-sense constructs. Crude protein extracts were prepared from each transformed plant and assayed using the method of Zhang et al. (Plant Physiol., 113:65-74, 1997). The data contained in the column labeled "Enzyme" in Table 2 shows that the transformed plant lines containing the 4CL sense construct had elevated 4CL enzyme activity, with a maximum of 258%, and the transformed plant lines containing the 4CL anti-sense construct had reduced 4CL enzyme activity, with a minimum of 59%, relative to empty vector-transformed control plants.

d) Effects of *Pinus* 4CL on lignin concentration in transformed plants

The concentration of lignin in samples of transformed plant material was determined as described in Example 3. The data shown in the column labelled "TGA" in Table 2 shows that the transformed plant lines containing the sense and the anti-sense 4CL gene constructs all exhibited significantly decreased levels of lignin, relative to the empty vector-transformed control plant lines. These data clearly indicate that

lignin concentration, as measured by the TGA assay, can be directly manipulated by either sense or anti-sense expression of a lignin biosynthetic gene such as 4CL.

Table 2

plant line	transgene	orientation	Southern	Northern	Enzyme	TGA
1	control	na	+	blank	100	92
2	control	na	+	blank	100	104
1	4CL	sense	+	2.3E+4	169	64
2	4CL	sense	+	4.5E+4	258	73
3	4CL	sense	+	3.1E+4	174	77
4	4CL	sense	+	1.7E+4	164	80
5	4CL	sense	+	1.6E+4	184	92
1	4CL	anti-sense	+	na	59	75
2	4CL	anti-sense	+	1.0E+4	70	75
3	4CL	anti-sense	+	9.6E+3	81	80
4	4CL	anti-sense	+	1.2E+4	90	83
5	4CL	anti-sense	+	4.7E+3	101	88
6	4CL	anti-sense	+	3.9E+3	116	89
7	4CL	anti-sense	+	1.8E+3	125	94
8	4CL	anti-sense	+	1.7E+4	106	97

Example 5

Transformation of Tobacco using the Inventive Lignin Biosynthetic Genes

Sense and anti-sense constructs containing sequences including the coding regions of C3H (SEQ ID NO: 18), F5H (SEQ ID NO: 19), CCR (SEQ ID NO: 25) and CGT (SEQ ID NO: 31) from *Eucalyptus grandis*, and PAL (SEQ ID NO: 45 and 47), C4H (SEQ ID NO: 48 and 49), PNL (SEQ ID NO: 81) and LAC (SEQ ID NO: 83) from *Pinus radiata* were inserted into *Agrobacterium tumefaciens* LBA4301 by direct transformation as described above. The presence and integrity of the transgenic constructs were verified by restriction digestion and DNA sequencing.

Tobacco (*Nicotiana tabacum* cv. Samsun) leaf sections were transformed as described in Example 3. Up to twelve independent transformed plant lines were established for each sense construct and each anti-sense construct listed in the preceding paragraph. Transformed plants containing the appropriate lignin gene

construct were verified using Southern blot experiments. All of the transformed plant lines analysed were confirmed as independent transformed lines.

Example 6

Manipulation of Lignin Content in Transformed Plants

a) Determination of transgene expression by Northern blot experiments

Total RNA was isolated from each independent transformed plant line described in Example 5. The RNA samples were analysed in Northern blot experiments to determine the level of expression of the transgene in each transformed line. The column labelled "Northern" in Table 3 shows the level of transgene expression for all plant lines assayed, relative to the background on the Northern blots. There was no detectable hybridisation to RNA samples from empty vector-transformed control plants.

b) Determination of lignin concentration in transformed plants

The concentration of lignin in empty vector-transformed control plant lines and in up to twelve independent transformed lines for each sense construct and each anti-sense construct described in Example 5 was determined as described in Example 3. The column labelled "TGA" in Table 3 shows the thioglycolic acid extractable lignins for all plant lines assayed, expressed as the average percentage of TGA extractable lignins in transformed plants versus control plants. The range of variation is shown in parentheses.

Table 3

transgene	orientation	no. of lines	Northern	TGA
control	na	3	blank	100 (92-104)
C3H	sense	5	3.7E+4	74 (67-85)
F5H	sense	10	5.8E+4	70 (63-79)
F5H	anti-sense	9	5.8E+4	73 (35-93)
CCR	sense	1	na	74
CCR	anti-sense	2	na	74 (62-86)
PAL	sense	5	1.9E+5	77 (71-86)
PAL	anti-sense	4	1.5E+4	62 (37-77)
C4H	anti-sense	10	5.8E+4	86 (52-113)
PNL	anti-sense	6	1.2E+4	88 (70-114)
LAC	sense	5	1.7E+5	na
LAC	anti-sense	12	1.7E+5	88 (73-114)

Transformed plant lines containing the sense and the anti-sense lignin biosynthetic gene constructs all exhibited significantly decreased levels of lignin, relative to the empty vector-transformed control plant lines. The most dramatic effects on lignin concentration were seen in the F5H anti-sense plants with as little as 35% of the amount of lignin in control plants, and in the PAL anti-sense plants with as little as 37% of the amount of lignin in control plants. These data clearly indicate that lignin concentration, as measured by the TGA assay, can be directly manipulated by conventional anti-sense methodology and also by sense over-expression using the inventive lignin biosynthetic genes.

Example 7

Modulation of Lignin Enzyme Activity in Transformed Plants

The activities and substrate specificities of selected lignin biosynthetic enzymes were assayed in crude extracts from transformed tobacco plants containing sense and anti-sense constructs for PAL (SEQ ID NO: 45), PNL (SEQ ID NO: 81) and LAC (SEQ ID NO: 83) from *Pinus radiata*, and CGT (SEQ ID NO: 31) from *Eucalyptus grandis*.

Enzyme assays were performed using published methods for PAL (Southerton, S.G. and Deverall, B.J., Plant Path. 39:223-230, 1990), CGT (Vellekoop, P. et al., FEBS, 330:36-40, 1993), PNL (Espin, C.J. et al., Phytochemistry, 44:17-22, 1997) and

LAC (Bao, W. et al., Science, 260:672-674, 1993). The data shown in the column labelled "Enzyme" in Table 4 shows the average enzyme activity from replicate measures for all plant lines assayed, expressed as a percent of enzyme activity in empty vector-transformed control plants. The range of variation is shown in parentheses.

Table 4

transgene	orientation	no. of lines	Enzyme
control	na	3	100
PAL	sense	5	87 (60-124)
PAL	anti-sense	3	53 (38-80)
CGT	anti-sense	1	89
PNL	anti-sense	6	144 (41-279)
LAC	sense	5	78 (16-240)
LAC	anti-sense	11	64 (14-106)

All of the transformed plant lines, except the PNL anti-sense transformed plant lines, showed average lignin enzyme activities which were significantly lower than the activities observed in empty vector-transformed control plants. The most dramatic effects on lignin enzyme activities were seen in the PAL anti-sense transformed plant lines in which all of the lines showed reduced PAL activity and in the LAC anti-sense transformed plant lines which showed as little as 14% of the LAC activity in empty vector-transformed control plant lines.

Example 8

Functional Identification of Lignin Biosynthetic Genes

Sense constructs containing sequences including the coding regions for PAL (SEQ ID NO: 47), OMT (SEQ ID NO: 53), 4CL (SEQ ID NO: 56 and 57) and POX (SEQ ID NO: 86) from *Pinus radiata*, and OMT (SEQ ID NO: 23 and 24), CCR (SEQ ID NO: 26-28), CGT (SEQ ID NO: 31 and 33) and POX (SEQ ID NO: 42 and 44) from *Eucalyptus grandis* were inserted into the commercially available protein expression vector, pProEX-1 (Gibco BRL). The resultant constructs were transformed into *E. coli* XL1-Blue (Stratagene), which were then induced to produce recombinant protein by the addition of IPTG. Purified proteins were produced for the *Pinus* OMT and 4CL constructs and the *Eucalyptus* OMT and POX constructs using Ni column

chromatography (Janknecht, R. et al., Proc. Natl. Acad. Sci., 88:8972-8976, 1991). Enzyme assays for each of the purified proteins conclusively demonstrated the expected substrate specificity and enzymatic activity for the genes tested.

The data for two representative enzyme assay experiments, demonstrating the verification of the enzymatic activity of a *Pinus radiata* 4CL gene (SEQ ID NO: 56) and a *Pinus radiata* OMT gene (SEQ ID NO: 53), are shown in Table 5. For the 4CL enzyme, one unit equals the quantity of protein required to convert the substrate into product at the rate of 0.1 absorbance units per minute. For the OMT enzyme, one unit equals the quantity of protein required to convert 1 pmole of substrate to product per minute.

Table 5

transgene	purification step	total ml extract	total mg protein	total units activity	% yield activity	fold purification
4CL	crude	10 ml	51 mg	4200	100	1
	Ni column	4 ml	0.84 mg	3680	88	53
OMT	crude	10 ml	74 mg	4600	100	1
	Ni column	4 ml	1.2 mg	4487	98	60

The data shown in Table 5 indicate that both the purified 4CL enzyme and the purified OMT enzyme show high activity in enzyme assays, confirming the identification of the 4CL and OMT genes described in this application. Crude protein preparations from *E. coli* transformed with empty vector show no activity in either the 4CL or the OMT enzyme assay.

Although the present invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, changes and modifications can be carried out without departing from the scope of the invention which is intended to be limited only by the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Genesis Research and Development Corp. Ltd.

(ii) TITLE OF THE INVENTION: MATERIALS AND METHODS FOR THE MODIFICATION OF PLANT LIGNIN CONTENT

(iii) NUMBER OF SEQUENCES: 88

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Russell McVeagh West-Walker
(B) STREET: The Todd Building, Cnr Brandon Street & Lambton Quay
(C) CITY: Wellington
(D) STATE:
(E) COUNTRY: New Zealand
(F) ZIP:

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: Wordperfect 5.1

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Bennett, Michael Roy
(B) REGISTRATION NUMBER:
(C) REFERENCE/DOCKET NUMBER: 22315\MRB

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: +64 4 495 7740
(B) TELEFAX: +64 4 499 9306
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 535 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTTCGCGCTA CCGCATACTC CACCACCGCG TGCAGAAGAT GAGCTCGGAG GGTGGGAAGG 60
AGGATTGCCT CGGTTGGGCT GCCCGGGACC CTTCTGGGTT CCTCTCCCCN TACAAATTCA 120
CCCGCAGGCC GTGGGAAGCG AAGACGTCTC GATTAAGATC ACGCACTGTG GAGTGTGCTA 180
CGCAGATGTG GCTTGGACTA GGAATGTGCA GGGACACTCC AAGTATCCTC TGGTGCCGGG 240

GCACGAGATA GTTGGAATTG TGAAACAGGT TGGCTCCAGT GTCCAACGCT TCAAAGTTGG
300
CGATCATGTG GGGGTGGGAA CTTATGTCAA TTCATGCAGA GAGTGCGAGT ATTGCAATGA
360
CAGGCTAGAA GTCCAATGTG AAAAGTCGGT TATGACTTTT GATGGAATTG ATGCAGATGG
420
TACAGTGACA AAGGGAGGAT ATTCTAGTCA CATTGTCGTC CATGAAAGGT ATTGCGTCAG
480
GATTCCAGAA AACTACCCGA TGGATCTAGC AGCGCATTGC TCTGTGCTGG ATCAC
535

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 671 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCGCCTGCAG GTCGACACTA GTGGATCCAA AGAATTCGGC ACGAGGTTGC AGGTCGGGGA
60
TGATTTGAAT CACAGAAACC TCAGCGATTT TGCCAAGAAA TATGGCAAAA TCTTTCTGCT
120
CAAGATGGGC CAGAGGAATC TTGTGGTAGT TTCATCTCCC GATCTCGCCA AGGAGGTCCT
180
GCACACCCAG GGCGTCGAGT TTGGGTCTCG AACCCGGAAC GTGGTGTTCG ATATCTTCAC
240
GGGCAAGGGG CAGGACATGG TGTTCACCGT CTATGGAGAT CACTGGAGAA AGATGCGCAG
300
GATCATGACT GTGCCTTTCT TTACGAATAA AGTTGTCCAG CACTACAGAT TCGCGTGGGA
360
AGACGAGATC AGCCGCGTGG TCGCGGATGT GAAATCCCGC GCCGAGTCTT CCACCTCGGG
420
CATTGTCATC CGTAGCGCCT CCAGCTCATG ATGTATAATA TTATGTATAG GATGATGTTC
480
GACAGGAGAT TCGAATCCGA GGACGACCCG CTTTTCCTCA AGCTCAAGGC CCTCAACGGA
540
GAGCGAAGTC GATTGGCCCA GAGCTTTGAG TACAATTATG GGGATTTCAT TCCCAGTCTT
600
AGGCCCTTCC TCAGAGGTTA TCACAGAATC TGCAATGAGA TTAAAGAGAA ACGGCTCTCT
660
CTTTTCAAGG A
671

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 940 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTTCAGGACA AGGGAGAGAT CAATGAGGAT AATGTTTTGT ACATCGTTGA GAACATCAAC
60
GTTGCAGCAA TTGAGACAAC GCTGTGGTCG ATGGAATGGG GAATAGCGGA GCTGGTGAAC
120
CACCAGGACA TTCAGAGCAA GGTGCGCGCA GAGCTGGACG CTGTTCTTGG ACCAGGCGTG
180
CAGATAACGG AACCAGACAC GACAAGGTTG CCCTACCTTC AGGCGGTTGT GAAGGAAACC
240

CTTCGTCTCC GCATGGCGAT CCCGTTGCTC GTCCCCCACA TGAATCTCCA CGACGCCAAG
300
CTCGGGGGCT ACGATATTCC GGCAGAGAGC AAGATCCTGG TGAACGCCTG GTGGTTGGCC
360
AACAACCCCG CCAACTGGAA GAACCCCGAG GAGTTCCGCC CCGAGCGGTT CTTCGAGGAG
420
GAGAAGCACA CCGAAGCCAA TGGCAACGAC TTCAAATTCC TGNCCTTCGG TGTGGGGAGG
480
AGGAGCTGCC CGGGAATCAT TCTGGCGCTG CTCTCCTCGC ACTCTCCATC GGAAGACTTG
540
TTCAGAACTT CCACCTTCTG CCGCCGCCCG GGCAGAGCAA AGTGGATGTC ACTGAGAAGG
600
GCGGGCAATT CAGCCTTCAC ATTCTCAACC ATTCTCTCAT CGTCGCCAAG CCCATAGCTT
660
CTGCTTAATC CCAACTTGTC AGTGACTGGT ATATAAATGC GCGCACCTGA ACAAAAAACA
720
CTCCATCTAT CATGACTGTG TGTGCGTGTC CACTGTCGAG TCTACTAAGA GCTCATAGCA
780
CTTCAAAAGT TTGCTAGGAT TTCAATAACA GACACCGTCA ATTATGTCAT GTTTCAATAA
840
AAGTTTGCAT AAATTAAATG ATATTTCAAT ATACTATTTT GACTCTCCAC CAATTGGGGA
900
ATTTTACTGC TAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
940

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 949 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

NNGCTCNACC GACGGTGGAC GGTCCGCTAC TCAGTAACTG AGTGGGATCC CCCGGGCTGA
60
CAGGCAATTC GATTTAGCTC ACTCATTAGG CACCCCAGGC TTTACACTTT ATGCTTCCGG
120
CTCGTATGTT GTGTGGAATT GTGAGCGGAT AACAATTTCA CACAGGAAAC AGCTATGACC
180
ATGATTACGC CAAGCGCGCA ATTAACCCTC ACTAAAGGGA ACAAAAGCTG GAGCTCCACC
240
GCGGTGGCGG CCGCTCTAGA ACTAGTGGAT CCAAAGAATT CGGCACGAGA CCCAGTGACC
300
TTCAGGCCTG AGAGATTTCT TGAGGAAGAT GTTGATATTA AGGGCCATGA TTACAGGCTA
360
CTGCCATTGG TGCAGGGCGC AGGATCTGCC CTGGTGCACA ATTGGGTATT AATTTAGTTC
420
AGTCTATGTT GGGACACCTG CTTCATCATT TCGTATGGGC ACCTCCTGAG GGAATGAAGG
480
CAGAAGACAT AGATCTCACA GAGAATCCAG GGCTTGTTAC TTTCATGGCC AAGCCTGTGC
540
AGGCCATTGC TATTCCTCGA TTGCCTGATC ATCTCTACAA GCGACAGCCA CTCAATTGAT
600
CAATTGATCT GATAGTAAGT TTGAATTTTG TTTTGATACA AAACGAAATA ACGTGCAGTT
660
TCTCCTTTTC CATAGTCAAC ATGCAGCTTT CTTTCTCTGA AGCGCATGCA GCTTTCTTTC
720
TCTGAAGCCC AACTTCTAGC AAGCAATAAC TGTATATTTT AGAACAAATA CCTATTCCTC
780
AAATTGAGWA TTTCTCTGTA GGGGNNGNTA ATTGTGCAAT TTGCAAGNAA TAGTAAAGTT
840
TANTTTAGGG NATTTTAATA GTCCTANGTA ANANGNGGNA ATGNTAGNGG GCATTNAGAA
900

ANCCCTAATA GNTGTTGGNG GNNGNTAGGN TTTTTNACCA AAAAAAAAA
949

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 959 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAATTCGGCA CGAGAAAGCC CTAGAATTTT TTCAGCATGC TATCACAGCC CCAGCGACAA
60
CTTTAACTGC AATAACTGTG GAAGCGTACA AAAAGTTTGT CCTAGTTTCT CTCATTCAGA
120
CTGGTCAGGT TCCAGCATTT CCAAAATACA CACCTGCTGT TGTCCAAAGA AATTTGAAAT
180
CTTGCACTCA GCCCTACATT GATTTAGCAA ACAACTACAG TAGTGGGAAA ATTTCTGTAT
240
TGGAAGCTTG TGTCAACACG AACACAGAGA AGTTCAAGAA TGATAGTAAT TTGGGGTTAG
300
TCAAGCAAGT TTTGTCATCT CTTTATAAAC GGAATATTCA GAGATTGACA CAGACATATC
360
TGACCCTCTC TCTTCAAGAC ATAGCAAGTA CGGTACAGTT GGAGACTGCT AAGCAGGCTG
420
AACTCCATGT TCTGCAGATG ATTCAAGATG GTGAGATTTT TGCAACCATA AATCAGAAAG
480
ATGGGATGGT GAGCTTCAAT GAGGATCCTG AACAGTACAA AACATGTCAG ATGACTGAAT
540
ATATAGATAC TGCAATTCGG AGAATCATGG CACTATCAAA GAAGCTCACC ACAGTAGATG
600
AGCAGATTTC GTGTGATCAT TCCTACCTGA GTAAGGTGGG GAGAGAGCGT TCAAGATTTG
660
ACATAGATGA TTTTGATACT GTTCCCCAGA AGTTCANAAA TATGTAACAA ATGATGTAAA
720
TCATCTTCAA GACTCGCTTA TATTCATTAC TTTCTATGTG AATTGATAGT CTGTTAACAA
780
TAGTACTGTG GCTGAGTCCA GAAAGGATCT CTCGGTATTA TCACTTGACA TGCCATCAAA
840
AAAATCTCAA ATTTCTCGAT GTCTAGTCTT GATTTTGATT ATGAATGCGA CTTTTAGTTG
900
TGACATTTGA GCACCTCGAG TGAACTACAA AGTTGCATGT TAAAAAAAAA AAAAAAAAA
959

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1026 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCGGCA CGAGCTTTGA GGCAACCTAC ATTCATTGAA TCCCAGGATT TCTTCTTGTC
60
CAAACAGGTT TAAGGAAATG GCAGGCACAA GTGTTGCTGC AGCAGAGGTG AAGGCTCAGA
120
CAACCCAAGC AGAGGAGCCG GTTAAGGTTG TCCGCCATCA AGAAGTGGGA CACAAAAGTC
180
TTTTGCAGAG CGATGCCCTC TATCAGTATA TATTGGAAAC GAGCGTGTAC CCTCGTGAGC
240

CCGAGCCAAT GAAGGAGCTC CGCGAAGTGA CTGCCAAGCA TCCCTGGAAC CTCATGACTA
300
CTTCTGCCGA TGAGGGTCAA TTTCTGGGCC TCCTGCTGAA GCTCATTAAC GCCAAGAACA
360
CCATGGAGAT TGGGGTGTAC ACTGGTTACT CGCTTCTCAG CACAGCCCTT GCATTGCCCG
420
ATGATGGAAA GATTCTAGCC ATGGACATCA ACAGAGAGAA CTATGATATC GGATTGCCTA
480
TTATTGAGAA AGCAGGAGTT GCCCACAAGA TTGACTTCAG AGAGGGCCCT GCTCTGCCAG
540
TTCTGGACGA ACTGCTTAAG AATGAGGACA TGCATGGATC GTTCGATTTT GTGTTCGTGG
600
ATGCGGACAA AGACAACTAT CTAAACTACC ACAAGCGTCT GATCGATCTG GTGAAGGTTG
660
GAGGTCTGAT TGCATATGAC AACACCCTGT GGAACGGATC TGTGGTGGCT CCACCCGATG
720
CTCCCCTGAG GAAATATGTG AGATATTACA GAGATTTCGT GATGGAGCTA AACAAGGCCC
780
TTGCTGTCGA TCCCCGCATT GAGATCAGCC AAATCCCAGT CGGTGACGGC GTCACCCTTT
840
GCAGGCGTGT CTATTGAAAA CAATCCTTGT TTCTGCTCGT CTATTGCAAG CATAAAGGCT
900
CTCTGATTAT AAGGAGAACG CTATAATATA TGGGGTTGAA GCCATTTGTT TTGTTTAGTG
960
TATTGATAAT AAAGTAGTAC AGCATATGCA AAGTTTGTAT CAAAAAAAAA AAAAAAAAAA
1020
AAAAAA
1026

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1454 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAATTCGGCA CGAGGCCAAC TGCAAGCAAT ACAGTACAAG AGCCAGACGA TCGAATCCTG
60
TGAAGTGGTT CTGAAGTGAT GGGAAGCTTG GAATCTGAAA AAACTGTTAC AGGATATGCA
120
GCTCGGGACT CCAGTGGCCA CTTGTCCCCT TACACTTACA ATCTCAGAAA GAAAGGACCT
180
GAGGATGTAA TTGTAAAGGT CATTTACTGC GGAATCTGCC ACTCTGATTT AGTTCAAATG
240
CGTAATGAAA TGGACATGTC TCATTACCCA ATGGTCCCTG GGCATGAAGT GGTGGGGATT
300
GTAACAGAGA TTGGCAGCGA GGTGAAGAAA TTCAAAGTGG GAGAGCATGT AGGGGTTGGT
360
TGCATTGTTG GGTCCTGTCG CAGTTGCGGT AATTGCAATC AGAGCATGGA ACAATACTGC
420
AGCAAGAGGA TTTGGACCTA CAATGATGTG AACCATGACG GCACACCTAC TCAGGGCGGA
480
TTTGCAAGCA GTATGGTGGT TGATCAGATG TWTGTGGTTC GAATCCCGGA GAATCTTCCT
540
CTGGAACAAG CGGCCCCTCT GTTATGTGCA GGGGTTACAG TTTTCAGCCC AATGAAGCAT
600
TTCGCCATGA CAGAGCCCGG GAAGAAATGT GGGATTTTGG GTTTAGGAGG CGTGGGGCAC
660
ATGGGTGTCA AGATTGCCAA AGCCTTTGGA CTCCACGTGA CGGTTATCAG TTCGTCTGAT
720
AAAAAGAAAG AAGAAGCCAT GGAAGTCCTC GGCGCCGATG CTTATCTTGT TAGCAAGGAT
780

ACTGAAAAGA TGATGGAAGC AGCAGAGAGC CTAGATTACA TAATGGACAC CATTCCAGTT
840
GCTCATCCTC TGGAACCATA TCTTGCCCTT CTGAAGACAA ATGGAAAGCT AGTGATGCTG
900
GGCGTTGTTC CAGAGTCGTT GCACTTCGTG ACTCCTCTCT TAATACTTGG GAGAAGGAGC
960
ATAGCTGGAA GTTTCATTGG CAGCATGGAG GAAACACAGG AAACTCTAGA TTTCTGTGCA
1020
GAGAAGAAGG TATCATCGAT GATTGAGGTT GTGGGCCTGG ACTACATCAA CACGGCCATG
1080
GAAAGGTTGG AGAAGAACGA TGTCCGTTAC AGATTTGTGG TGGATGTTGC TAGAAGCAAG
1140
TTGGATAATT AGTCTGCAAT CAATCAATCA GATCAATGCC TGCATGCAAG ATGAATAGAT
1200
CTGGACTAGT AGCTTAACAT GAAAGGGAAA TTAAATTTTT ATTTAGGAAC TCGATACTGG
1260
TTTTTGTTAC TTTAGTTTAG CTTTTGTGAG GTTGAAACAA TTCAGATGTT TTTTTAACTT
1320
GTATATGTAA AGATCAATTT CTCGTGACAG TAAATAATAA TCCAATGTCT TCTGCCAAAT
1380
TAATATATGT ATTCGTATTT TTATATGAAA AAAAAAAAAA AAAA
1454AAAAAA AAAAAAAAAA 1440 AAAAAAAAAA AAAA
1454

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 740 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAATTCGGCA CGAGACCATT TCCAGCTAAT ATTGGCATAG CAATTGGTCA TTCTATCTTT
60
GTCAAAGGAG ATCAAACAAA TTTTGAAATT GGACCTAATG GTGTGGAGGC TAGTCAGCTA
120
TACCCAGATG TGAAATATAC CACTGTCGAT GAGTACCTCA GCAAATTTGT GTGAAGTATG
180
CGAGATTCTC TTCCACATGC TTCAGAGATA CATAACAGTT TCAATCAATG TTTGTCCTAG
240
GCATTTGCCA AATTGTGGGT TATAATCCTT CGTAGGTGTT TGGCAGAACA GAACCTCCTG
300
TTTAGTATAG TATGACGAGC TAGGCACTGC AGATCCTTCA CACTTTTCTC TTCCATAAGA
360
AACAAATACT CACCTGTGGT TTGTTTTCTT TCTTTCTGGA ACTTTGGTAT GGCAATAATG
420
TCTTTGGAAA CCGCTTAGTG TGGAATGCTA AGTACTAGTG TCCAGAGTTC TAAGGGAGTT
480
CCAAAATCAT GGCTGATGTG AACTGGTTGT TCCAGAGGGT GTTTACAACC AACAGTTGTT
540
CAGTGAATAA TTTTGTTAGA GTGTTTAGAT CCATCTTTAC AAGGCTATTG AGTAAGGTTG
600
GTGTTAGTGA ACGGAATGAT GTCAAATCTT GATGGGCTGA CTGACTCTCT TGTGATGTCA
660
AATCTTGATG GATTGTGTCT TTTTCAATGG TAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
740 720 AAAAAAAAAA AAAAAAAAAA
740

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 624 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAATTCCTGC AGCCCGGGGG ATCCACTAGT TCTAGAGCGG CCGCCACCGC GGTGGAGCTC
60
GCGCGCCTGC AGGTCGACAC TAGTGGATCC AAAGAATTCG GCACGAGGCC CGACGGCCAC
120
TTGTTGGACG CCATGGAAGC TCTCCGGAAA GCCGGGATTC TGGAACCGTT TAAACTGCAG
180
CCCAAGGAAG GACTGGCTCT CGTCAACGGC ACAGCGGTGG GATCCGCCGT GGCCGCGTCC
240
GTCTGTGTTG ACGCCAACGT GCTGGGCGTG CTGGCTGAGA TTCTGTCTGC GCTCTTCTGC
300
GAGGTGATGC AAGGGAAACC GGAGTTCGTA GATCCGTTAA CCCACCAGTT GAAGCACCAC
360
CCAGGGCAGA TCGAAGCCGC GGCCGTCATG GAGTTCCTCC TCGACGGTAG CGACTACGTG
420
AAAGAAGCAG CGCGGCTTCA CGAGAAAGAC CCGTTGAGCA AACCGAAACA AGACCGCTAC
480
GCTCTGCGAA CATCGCCACA GTGGTTGGGG CCTCCGATCG AAGTCATCCG CGCTGCYACT
540
CACTCCATCG AGCGGGAGAT CAATTCCGTC AACGACAATC CGTTAATCGA TGTCTCCAGG
600
GACATGGCTG TCCACGGCGG CAAC
624

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAATTCCTGC AGCCCGGGGG ATCCACTAGT TCTAGAGCGG CCGCCACCGC GGTGGAGCTC
60
CAGTACCTGG CCAACCCCGT CACGACTCAC GTCCAGAGCG CCGAACAACA CAACCAGGAT
120
GTCAATTCCC TCGGCTTGAT CTCCGCCAGA AAGACTGCCG AGGCCGTTGA GATTTTAAAG
180
CTGATGTTCG CTACATATCT GGTGGCCTTA TGCCAGGCGA TCGATCTCCG GCACCTGGAA
240
GAAAACATGC GATCCGTTGT GAAGCACGTA GTCTTGCA
278

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 765 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAGCTCCTGC AAGTCATCGA TCATCAGCCC GTTTTCTCGT ACATCGACGA TCCCACAAAT
60
CCATCATACG CGCTTATGCT CCAACTCAGA GAAGTGCTCG TAGATGAGGC TCTCAAATCA
120

TCTTGCCCAG ACGGGAATGA CGAATCCGAT CACAATTTGC AGCCCGCTGA GAGCGCTGGA
180
GCTGCTGGAA TATTACCCAA TTGGGTGTTT AGCAGGATCC CCATATTTCA AGAGGAGTTG
240
AAGGCCCGTT TAGAGGAAGA GGTTCCGAAG GCGAGGGAAC GATTCGATAA TGGGGACTTC
300
CCAATTGCAA ACAGAATAAA CAAGTGCAGG ACATATCCCA TTTACAGATT CGTGAGATCA
360
GAGTTGGGAA CCGATTTGCT AACAGGGCCC AAGTGGAGAA GCCCCGGCGA AGATATAGAA
420
AAGGTATTTG AGGGCATTTG CCAAGGGAAA ATTGGAAACG TGATCCTCAA ATGTCTGGAC
480
GCTTGGGGTG GGTGCGCTGG ACCATTCACT CCACGTGCAT ATCCTGCGTC TCCTGCAGCG
540
TTCAATGCCT CATATTGGGC ATGGTTTGAT AGCACCAAAT CACCCTCTGC AACGAGCGGC
600
AGAGGTTTCT GGAGCGCCCA ACAACAACAA GTTCTTTGAT TTAACTGACT CTTAAGCATT
660
CCTAAACAGC TTGTTCTTCG CAATAACGAA TCTTTCATCT TCGTTACTTT GTAAAAGATG
720
GGGTTCCAAC AAAATAGAAG AAATATTTTC GATCCAAAAA AAAAA
765

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 453 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TGATTATGCG GATCCTTGGG CAGGGATACG GCATGACAGA AGCAGGCCCG GTGCTGGCAA
60
TGAACCTAGC CTTCGCAAAG AATCCTTTCC CCGCCAAATC TGGCTCCTGC GGAACAGTCG
120
TCCGGAACGC TCAAATAAAG ATCCTCGATT ACAGGAACTG GCGAGTCTCT CCCGCACAAT
180
CAAGCCGGCG AAATCTGCAT CCGCGGACCC GAAATAATGA AAGGATATAT TAACGACCCG
240
GAATCCACGG CCGCTACAAT CGATGAAGAA GGCTGGCTCC ACACAGGCGA CGTCGGGTAC
300
ATTGACGATG ACGAAGAAAT CTTCATAGTC GACAGAGTAA AGGAGATTAT CAATATAAAG
360
GCTTCCAGGT GGATCCTGCT AATCGAATTC CTGCAGCCCG GGGGTCCACT AGTTCTAGAG
420
CGGCCGCCAC CGCGGTGGAG CTCCAGCTTT TGT
453

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TCTTCGAATT CTCTTTCACG ACTGCTTCGT TAATGGCTGC GATGGCTCGA TATTGTTAGA
60
TGATAACTCA ACGTTCACCG GAGAAAAGAC TGCAGGCCCA AATGTTAATT CTGCGAGAGG
120

ATTCGACGTA ATAGACACCA TCAAAACTCA AGTTGAGGCA GCCTGCAGTG GTGTCGTGTC
180
AGTTGCCGAC ATTCTCGCCA TTGCTGCACG CGATTCAGTC GTCCAACTGG GGGGCCCAAC
240
ATGGACGGTA CTTCTGGGAG AAAAGACGGA TCCGATCA
278

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTTCGAATTC WYTTYCAYGA YTG
23

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GATCGGATCC RTCYYKYCTY CC
22

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 472 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AATTCGGCAC GAGACGACCT CTTGTATCGG ACCCGGATCC GCTATCGTTA ACGTACACAC
60
GTTCTAGTGC TGAATGGAGA TGGAGAGCAC CACCGGCACC GGCAACGGCC TTCACAGCCT
120
CTGCGCCGCC GGGAGCCACC ATGCCGACCC ACTGAACTGG GGGGCGGCGG CAGCAGCCCT
180
CACAGGGAGC CACCTCGACG AGGTGAAGCG GATGGTCGAG GAGTACCGGA GGCCGGCGGT
240
GCGCCTCGGC GGGGAGTCCC TCACGATAGC CCAGGTGGCG GCGGTGGCGA GTCAGGAGGG
300
GGTAGGGGTC GAGCTCTCGG AGGCGGCCCG TCCCAGGGTC AAGGCCAGCA GCGACTGGGT
360
CATGGAGAGC ATGAACAAGG GAACTGACAG CTACGGGGTC ACCACCGGGT TCGGCGGCAA
420
CTTCTCAAAC CGGAGGCCGA AGCAAGGCGG TCCTTTTCAG AAGGAACTTA TA
472

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCAAAGCTCC TAGTGCCTCA TGAGTCTGCT GAGGATTGCA CAATTGGCGG GTTCGACGTG
60
CCCCGAGGCA CCATGATCCT GGTTAATGCG TGGGCAATTC AAAGAGACCC AAAAGTGTGG
120
GACGATCCCA CAAATTTTAA ACCGGAGAGG TACGAGGGAT TGGAAGGTGA TCATGCCTAC
180
CGACTATTGC CGTTTGGGAT GGGGAGGAGA AGTTGTCCTG GTGCTGGCCT TGCCAATAGA
240
GTGGTGAGCT TGGTCCTGGC GGCGCTTATT CAGTGCTTCG AATGGGAACG AGTTGGCGAA
300
GAATTGGTGG ACTTGTCCGA GGGGACGGGA CTCACAATGC CAAAGAGAGA GCCATTGGAG
360
GCCTTGTGCA AAGCGCGTGA ATGCATGATA GCTAATGTTC TTGCGCACCT TTAAGAAGGT
420
CGTTGTCTAA TGAATTTACA TTGGTGATGT ATCTCCAATG TTTTTGAATA ATCAAATAGA
480
CTGAAAATAG GCCAGTGCAG CTTTAGGAAT GATCGTGAGC ATCAATAGCA TCCTGAGGAG
540
GCCAATGCAG CTTTAGGCCT TTCTCTTAGG AGAAAAATGA TGGTTTATAT AGGTACTGGC
600
AACATTGTTC AAAAAAAAAA AA
622

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 414 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CACGCTCGAC GAATTCGGTA CCCCGGGTTC GAAATCGATA AGCTTGGATC CAAAGCAACA
60
CATTGAACTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCCCCCACCC CCCCTTCCCA
120
ACCCCACCCA CATACAGACA AGTAGATACG CGCACACAGA AGAAGAAAAG ATGGGGGTTT
180
CAATGCAGTC AATCGCACTA GCGACGGTTC TGGCCGTCCT AACGACATGG GCGTGGAGGG
240
CGGTGAACTG GGTGTGGCTG AGGCCGAAGA GGCTCGAGAG GCTTCTGAGA CAGCAAGGTC
300
TCTCCGGCAA GTCCTACACC TTCCTGGTCG GCGACCTCAA GGAGAACCTG CGGATGCTCA
360
AGGAAGCCAA GTCCAAGCCC ATCGCCGTCT CCGATGACAT CAAGCCTCGT CTCT
414

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 469 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATTCGGCA CGAGTGTCTC TCTCTCTCTC TCTCTCTGTA AACCACCATG CTCTTCCTCA
60
CTCATCTCCT AGCAGTTCTA GGGGTTGTGT TGCTCCTGCT AATTCTATGG AGGGCAAGAT
120
CTTCTCCGAA CAAACCCAAA GGTACTGCCT TACCCCCGGA GCTGCCGGGC GCATGGCCGA
180
TCATAGGCCA CATCCACTTG CTGGGCGGCG AGACCCCGCT GGCCAGGACC CTGGCCGCCA
240
TGGCGGACAA GCAGGGCCCG ATGTTTCGGA TCCGTCTCGG AGTCCACCCG GCGACCATCA
300
TAAGCAGCCG TGAGGCGGTC CGGGAGTGCT TCACCACCCA CGACAAGGAC CTCGCTTCTC
360
GCCCCAAATC CAAGGCGGGA ATCCACTTGG GCTACGGGTA TGCCGGTTTT GGCTTCGTAG
420
AATACGGGGA CTTTTGGCGC GAGATGAGGA AGATCACCAT GCTCGAGCT
469

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 341 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGGGCTCGTG GCTCGGCTCC GGCGCAACGC CCTTCCCACC GGGCCCGAGG GGCCTCCCGG
60
TCATCGGGAA CATGCTCATG ATGGGCGAGC TCACCCACCG CGGCCTCGCG AGTCTGGCGA
120
AGAAGTATGG CGGGATCTTC CACCTCCGCA TGGGCTTCCT GCACATGGTT GCCGTGTCGT
180
CCCCCGACGT GGCCCGCCAG GTCCTCCAGG TCCACGACGG GATCTTCTCG AACCGGCCTG
240
CCACCATCGC GATCAGCTAC CTCACGTATG ACCGGGCCGA CATGGCCTTC GCGCACTACG
300
GCCCGTTCTG GCGGCAGATG CGGAAGCTGT GCGTGATGAA A
341

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 387 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GAATTCGGCA CGAGCGGGCT CGTGGCTCGG CTCCGGCGCA ACGCCCTTCC CACCGGGCCC
60
GAGGGGCCTC CCGGTCATCG GGAACATGCT CATGATGGGC GAGCTCACCC ACCGCGGCCT
120
CGCGAGTCTG GCGAAGAAGT ATGGCGGGAT CTTCCACCTC CGCATGGGCT TCCTGCACAT
180
GGTTGCCGTG TCGTCCCCCG ACGTGGCCCG CCAGGTCCTC CAGGTCCACG ACGGGATCTT
240
CTCGAACCGG CCTGCCACCA TCGCGATCAG CTACCTCACG TATGACCGGG CCGACATGGC
300
CTTCGCGCAC TACGGCCCGT TCTGGCGGCA GATGCGGAAG CTGTGCGTGA TGAAAGCTCT
360

TCAGCGGAAG CGGGCTGAGT CGTGGGA
387

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CACGAGCTCG TGAGCCTTCC CGGAGACAAG GCCATCTTAC TTCGCAACAA ATTGCGTCCG
60
CACTCCTTTC TCAAGAAACC TAGTCATCCA AGAAGCAGAG CATTGCAACT GCAAACAGCC
120
AAAGCCCAAA CTCGTACAGA AGGAGAGAGA GAGAGAGAAT AGAAGCATGA GTGCATGCAC
180
GAACCAAGCA ATCACGACGG CCAGTGAAGA TGAAGAGTTC TTGTTCGCCA TGGAAATGAA
240
TGCTCTGATA GCACTCCCCT TGGTCTTGAA GGCCACCATC GAACTGGGGA TCCTCGAAAT
300
ACTGGCCGAG TGCGGGCCTA TGGCTCCACT TTCGCCTGCT CAGATTGCCT CCCGTCTCTC
360
CGCAAAGAAC CCGGAAGCCC CCGTAACCCT TGACCGGATC CTCCGGTTTC TCGCCAGCTA
420
CTCCATCCTC TCTTGCACTC TCG
443

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 607 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GAATTCGGCA CGAGCCAACC CTGGACCAGG TACTTTTGGC AGGCGGTCCA TTGCCCTTCA
60
AACCGGTCCA AACCGGACCA TCACTGTCCT TATATACGTT GCATCATGCC TGCTCATAGA
120
ACTTAGGTCA ACTGCAACAT TTCTTGATCA CAACATATTA CAATATTCCT AAGCAGAGAG
180
AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGTTTGAA TCAATGGCCA CCGCCGGAGA
240
GGAGAGCCAG ACCCAAGCCG GGAGGCACCA GGAGGTTGGC CACAAGTCTC TCCTTCAGAG
300
TGATGCTCTT TACCAATATA TTTTGGAGAC CAGCGTGTAC CCAAGAGAGC CTGAGCCCAT
360
GAAGGAGCTC AGGGAAATAA CAGCAAAACA TCCATGGAAC ATAATGACAA CATCAGCAGA
420
CGAAGGGCAG TTCTTGAACA TGCTTCTCAA GCTCATCAAA GCCAAGAACA CCATGGAGAT
480
TGGTGTCTTC ACTGGCTACT CTCTCCTCGC CACCGCTCTT GCTCTTCCTG ATGACGGAAA
540
GATTTTGGCT ATGGACATTA ACAGAGAGAG CTATGAACTT GGCCTGCCGG CATCCAAAAA
600
GCCGGTG
607

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 421 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAATTCGGCA CGAGCCGTTT TATTTCCTCT GATTTCCTTT GCTCGAGTCT CGCGGAAGAG
60
AGAGAAGAGA GGAGAGGAGA GAATGGGTTC GACCGGATCC GAGACCCAGA TGACCCCGAC
120
CCAAGTCTCG GACGAGGAGG CGAACCTCTT CGCCATGCAG CTGGCGAGCG CCTCCGTGCT
180
CCCCATGGTC CTCAAGGCCG CCATCGAGCT CGACCTCCTC GAGATCATGG CCAAGGCCGG
240
GCCGGGCGCG TTCCTCTCCC CGGGGGAAGT CGCGGCCCAG CTCCCGACCC AGAACCCCGA
300
GGCACCCGTA ATGCTCGACC GGATCTTCCG GCTGCTGGCC AGCTACTCCG TGCTCACGTG
360
CACCCTCCGC GACCTCCCCG ATGGCAAGGT CGAGCGGCTC TACGGCTTAG CGCCGGTGTG
420
C
421

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 760 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGAAGAAGCC GAGCAAACGA ATTGCAGACG CCATTGAAAA AAGACACGAA AGAGATCAAG
60
AAGGAGCTTA AGAAGCATCA TCAATGGCAG CCAACGCAGA GCCTCAGCAG ACCCAACCAG
120
CGAAGCATTC GGAAGTCGGC CACAAGAGCC TCTTGCAGAG CGATGCTCTC TACCAGTATA
180
TATTGGAGAC CAGCGTCTAC CCAAGAGAGC CAGAGCCCAT GAAGGAGCTC AGGGAAATAA
240
CAGCCAAACA TCCATGGAAC CTGATGACCA CATCGGCGGA TGAAGGGCAG TTCCTGAACA
300
TGCTCCTCAA GCTCATCAAC GCCAAGAACA CCATGGAGAT CGGCGTCTAC ACCGGCTACT
360
CTCTCCTCGC AACCGCCCTT GCTCTTCCCG ATGACGGAAA GATCTTGGCC ATGGCCATCA
420
ATAGGGAGAA CTTCGAGATC GGGCTGCCCG TCATCCAGAA GGCCGGCCTT GCCCACAAGA
480
TCGATTTCAG AGAAGGCCCT GCCCTGCCGC TCCTTGATCA GCTCGTGCAA GATGAGAAGA
540
ACCATGGAAC GTACGACTTC TTCTCAATCC TTAATCGTTC ATTTGAATAC AAATACATGC
600
TCAATGGTTC AAAGACAACA TAAGACAGAA GATGGAAAAA ATAGAAAGGA AGGAAAGTAT
660
TAAGGGTAGT TTCTCATTTC ATCAATGCTT GATTTTGAGA TCTCCTTTCT GGTGCGATCA
720
GCTGACCCGG CGGCACAGGT GATGCCATCC CCGACGGGAA
760

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 508 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GAATTCGGTA CCCGGGTTCG AAATCGATAA GCTTGGATCC AAAGAATTCG GCACGAGATC
60
ACTAACCATC TGCCTTTCTT CATCTTCTTT CTTCTGCTTC TCCTCCGTTT CCTCGTTTCG
120
ATATCGTGAA AGGAGTCCGT CGACGACAAT GGCCGAGAAG AGCAAGGTCC TGATCATCGG
180
AGGGACGGGC TACGTCGGCA AGTTCATCGT GGAAGCGAGT GCAAAAGCAG GGCATCCCAC
240
GTTCGCGCTG GTTAGGCAGA GCACGGTCTC CGACCCCGTC AAGGGCCAGC TCGTCGAGAG
300
CTTCAAGAAC TTGGGCGTCA CTCTGCTCAT CGGTGATCTG TACGATCATG AGAGCTTGGT
360
GAAGGCAATC AAGCAAGCCG ACGTGGTGAT ATCGACAGTG GGGCACATGC AAATGGCGGA
420
TCAGACCAAA GAATCGTCGA CGCCATTAAA GGAAGCTGGC AACGTTAAGG TTTGTTGGTT
480
GGTTCATTTG ATCTGGTTTG GGGGGGTC
508

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 495 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GAATTCGGCA CGAGGTTAAT GGCAGTGCAG CCTCAACACC ACCCACCTTC CTCCATCTCT
60
CTCCTCCCTT CTTCTTTCTC TGACTTCAAT GGCAGCCGAC TCCATGCTTG CGTTCAGTAT
120
AAGAGGAAGG TGGGGCAGCC TAAAGGGGCA CTGCGGGTCA CTGCATCAAG CAATAAGAAG
180
ATCCTCATCA TGGGAGGCAC CCGTTTCATC GGTGTGTTTT TGTCGAGACT ACTTGTCAAA
240
GAAGGTCATC AGGTCACTTT GTTTACCAGA GGAAAAGCAC CCATCACTCA ACAATTGCCT
300
GGTGAGTCGG ACAAGGACTT CGCTGATTTT TCATCCAAGA TCCTGCATTT GAAAGGAGAC
360
AGAAAGGATT TTGATTTTGT TAAATCTAGT CTTGCTGCAG AAGGCTTTGA CGTTGTTTAT
420
GACATTAACG GCGAGAGGCG GATGAAGTCG CACCAATTTT GGATGCCTGC CAAACCTTGA
480
ACCAGTCAAC TACTG
495

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 472 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GAATTCGGCA CGAGCATAAG CTCTCCCGTA ATCCTCACAT CACATGGCGA AGAGCAAGGT
60
CCTCGTCGTT GGCGGCACTG GCTACCTCGG GCGGAGGTTC GTGAGGGCGA GCCTGGACCA
120
GGGCCACCCC ACGTACGTCC TCCAGCGTCC GGAGACCGGC CTCGACATTG AGAAGCTCCA
180
GACGCTACTG CGCTTCAAGA GGCGTGGCGC CCAACTCGTC GAGGCCTCGT TCTCAGACCT
240
GAGGAGCCTC GTCGACGCTG TGAGGCGGGT CGATGTCGTC GTCTGTGCCA TGTCGGGGGT
300
CCACTTCCGG AGCCACAACA TCCTGATGCA GCTCAAGCTC GTGGAGGCTA TCAAAGAAGC
360
TGGAAATGTC AAGCGGTTTT TGCCGTCAGA GTTCGGAATG GACCCGGCCC TCATGGGTCA
420
TGCAATTGAG CCGGGAAGGG TCACGTTCGA TGAGAAATGG AGGTGAGAAA AG
472

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GAATTCGGCA CGAGGAGGCA CCTCCTCGAA ACGAAGAAGA AGAAGGACGA AGGACGAAGG
60
AGACGAAGGC GAGAATGAGC GCGGCGGGCG GTGCCGGGAA GGTCGTGTGC GTGACCGGGG
120
CGTCCGGTTA CATCGCCTCG TGGCTCGTCA AGCTCCTCCT CCAGCGCGGC TACACCGTCA
180
AGGCCACCGT CCGCGATCCG AATGATCCAA AAAAGACTGA ACATTTGCTT GGACTTGATG
240
GAGCGAAAGA TAGACTTCAA CTGTTCAAAG CAAACCTGCT GGAAGAGGGT TCATTTGATC
300
CTATTGTTGA GGGTTGTGCA GGCGTTTTTC AAACTGCCTC TCCCTTTTAT CATGATGTCA
360
AGGATCCGCA GGCAGAATTA CTTGATCCGG CTGTAA
396

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 592 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GAATTCGGCA CGAGGTTGAA CCTCCCGTCC TCGGCTCTGC TCGGCTCGTC ACCCTCTTCG
60
CGCTCCCGCA TACTCCACCA CCGCGTACAG AAGATGAGCT CGGAGGGTGG GAAGGAGGAT
120
TGCCTCGGTT GGGCTGCCCG GGACCCTTCT GGGTTCCTCT CCCCCTACAA ATTCACCCGC
180
AGGGCCCTGG GAAGCGAAGA CGTCTCGATT AAGATCACGC ACTGTGGAGT GTGCTACGCA
240

GATGTGGCTT GGACTAGGAA TGTGCAGGGA CACTCCAAGT ATCCTCTGGT GCCAGGGCAC
300
GAGATAGTTG GAATTGTGAA ACAGGTTGGC TCCAGTGTCC AACGCTTCAA AGTTGGCGAT
360
CATGTGGGGG TGGGAACTTA TGTCAATTCA TGCAGAGAGT GCGAGTATTG CAATGACAGG
420
CTAGAAGTCC AATGTGAAAA GTCGGTTATG ACTTTTGATG GAATTGATGC AGATGGTACA
480
GTGACAAAGG GAGGATATTC TAGTCACATT GTCGTCCATG AAAGGTATTG CGTCAGGATT
540
CCAGAAAACT ACCCGATGGA TCTAGCAGCG CATTTGCTCT GTGCTGGATC AC
592

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 468 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GAATTCGGCA CGAGAACTCA TCTTGAAATG TCATTGGAGT CATCATCCTC TAGTGAGAAG
60
AAACAAATGG GTTCCGCCGG ATTCGAATCG GCCACAAAGC CGCACGCCGT TTGCATTCCC
120
TACCCTGCAC AAAGCCACAT TGGCGCCATG CTCAAGCTAG CAAAGCTCCT CCATCACAAG
180
GGCTTCCACA TCTCCTTCGT CAACACCGAG TTCAACCACC GGCGGCTCGC CAGGGCTCGA
240
GGCCCCGAGT TCACAAATGG AATGCTGAGC GACTTTCAGT TCCTGACAAT CCCCGATGGT
300
CTTCCTCCTT CGGACTTGGA TGCGATCCAA GACATCAAGA TGCTCTGCGA ATCGTCCAGG
360
AACTATATGG TCAGCCCCAT CAACGATCTT GTATCGAGCC TGGGCTCGAA CCCGAGCGTC
420
CCTCCGGTGA CTTGCATCAA TCTCGGATGG TTTCATGACA CTCGTGAC
468

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 405 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CTTTACTCCG CCAAGAAGAT CCAATCGCAG TTTTCGCAAT TGGCCCATTA CACAAATGCG
60
GTCCATCTTC ATCGGGAAGT CTCTTGGCAG AAGACCGGAG TTGCATTTCC TGGCTGGACA
120
AGCAAGCCCC TAACTCAGTG GTCTATGTGA GTCTTGGGAG CATCGCCTCT GTGAACGAGT
180
CGGAATTTTC CGAAATAGCT TTAGGTTTAG CCGATAGCCA GCAGCCATTC TTGTGGGTGG
240
TTCGACCCGG GTCAGTGAGC GGCTCGGAAC TCTTAGAGAA TTTGCCCGGT TGCTTTCTGG
300
AGGCATTACA GGAGAGGGGG AAGATTGTGA AATGGGCGCC TCAACATGAA GTGCTGGCTC
360
ATCGGGCTGT CGGAGCGTTT TGGACTCACA ATGGATGGAA CTCCA
405

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 380 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGCAAACACG CCCGTTTTCG TTTTACTAAG AGAAGATGGT GAGCGTTGTG GCTGGTAGAG
60
TCGAGAGCTT GTCGAGCAGT GGCATTCAGT CGATCCCGCA GGAGTATGTG AGGCCGAAGG
120
AGGAGCTCAC AAGCATTGGC GACATCTTCG AGGAGGAGAA GAAGCATGAG GGCCCTCAGG
180
TCCCGACCAT CGACCTCGAG GACATAGCGT CTAAAGACCC CGTGGTGAGG GAGAGGTGCC
240
ACGAGGAGCT CAGGAAGGCT GCCACCGACT GGGGCGTCAT GCACCTCGTC AACCATGGGA
300
TCCCCAACGA CCTGATTGAG CGTGTAAAGA AGGCTGGCGA GGTGTTCTTC AACCTCCCGA
360
TCGAGGAGAA GGACAAGCAT
380

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 305 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TTGTACCCGA AGATCTCCGG GACCGTTCGA CGGCGACATC GCCGTCGGCC GGGAACCCGT
60
CGAGGCCGCC GCCGGAGGCC GGGGAGAAGC TGGAGTAGCC GCCGTAGCCG GAGAAGGCGC
120
CGTCGTGGTC GGCGGCGGCG GCGTGGTGGA CCTCATCGCC GTCCATGCTG AAGGCGTCGA
180
AGGAAGCGGA CATGGCTGGG GGATCGATCG ACCGATCCGA TCGGCCGGAG GATTTCGAGA
240
TCGGAGATGG AGAGATGGAA ATGAAAGAGA GAGAGAGAGA GAGATCCGGT GGACTGGTGG
300
TGTTT
305

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 693 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GAATTCGGCA CGAGCTAAGA GAGGAGAGGA GAGGAGCAAG ATGGCACTAG CAGGAGCTGC
60
ACTGTCAGGA ACCGTGGTGA GCTCCCCCTT TGTGAGGATG CAGCCTGTGA ACAGACTCAG
120

GGCATTCCCC AATGTGGGTC AGGCCCTGTT TGGTGTCAAC TCTGGCCGTG GCAGAGTGAC
180
TGCCATGGCC GCTTACAAGG TCACCCTGCT CACCCCTGAA GGCAAAGTCG AACTCGACGT
240
CCCCGACGAT GTTTACATCT TGGACTACGC CGAGGAGCAA GGCATCGACT TGCCCTACTC
300
CTGCCGTGCC GGCTCTTGCT CCTCCTGCGC GGGCAAGGTC GTGGCGGGGA GCGTCGACCA
360
GAGCGACGGC AGCTTCCTGG ATGATGATCA GATTGAGGAA GGTTGGGTCC TCACTTGTGT
420
CGCCTACCCT AAGTCTGAGG TCACCATTGA GACCCACAAG GAAGAGGAGC TCACTGCTTG
480
AAGCTCTCCT ATATTTGCTT TTGCATAAAT CAGTCTCACT CTACGCAACT TTCTCCACTC
540
TCTCCCCCCT TCACTACATG TTTGTTAGTT CCTTTAGTCT CTTCCTTTTT TACTGTACGA
600
GGGATGATTT GATGTTATTC TGAGTCTAAT GTAATGGCTT TTCTTTTTCC TATTTCTGTA
660
TGAGGAAATA AAACTCATGC TCTAAAAAAA AAA
693

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 418 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

AGGACTTTAT TATAAGCATT GTAAAAAGAG TCAAACTAAT ACATCGCAAG AATTGGGTTA
60
TCCAATAATC TACAAAAAGA AAAAAGTTTG ATGCATTGAG ATGGTAACTG CTTAATTCAA
120
ATGCCTTAGT TTGAAAAATT AACCAACTAT TAAAATTAAT GATGATGAAT ATGGATTATG
180
TGTGAAAAAC TATATAGACT TAAAATTGAC TCAGAAGACA TTCTTTTCTT CTTATTTTAT
240
GATATGATGA ATTCGGTCTA AACAGGCAAA TGGTGTCAAA CGGGAAGTCG GCAAAACTCT
300
TCCTCGGCAG TGACTACCGG GCGGGCGATG ATGCGGATCC GGGGGCCGGG TCGCTGGAGA
360
ACATCCCGCA CGGACCGGTC CACGTTTGGT GCGGTGACAA CAGGCAGCCC AACCTGGA
418

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 777 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GAATTCGGCA CGAGCATACA ACTACACTGC GACGCCGCCG CAGAACGCGA GCGTGCCGAC
60
CATGAACGGC ACCAAGGTCT ACCGGTTGCC GTATAACGCT ACGGTCCAGC TCGTTTTACA
120
GGACACCGGG ATAATCGCGC CGGAGACCCA CCCCATCCAT CTGCACGGAT TCAACTTCTT
180
CGGTGTGGGC AAAGGAGTGG GGAATTATGA CCCAAAGAAG GATCCCAAGA AGTTCAATCT
240

GGTTGACCCA GTGGAGAGGA ACACCATTGG AATCCCATCT GGTGGATGGA TAGCCATCAG
300
ATTCACAGCA GACAATCCAG GAGTTTGGTT CCTGCACTGC CATCTGGAAG TGCACACAAC
360
TTGGGGACTG AAGATGGCAT TCTTGGTGGA CAATGGGAAG GGGCCTAAAG AGACCCTGCT
420
TCCACCTCCA AGTGATCTTC CAAAATGTTG ATCATTTGAT CATGAGGACG ACAAGCGATT
480
ACTAATGACA CCAAGTTAGT GGAATCTTCT CTTTGAAAAA GAAGAAGAAG AGCAAGAAGA
540
ATAAGAAAGA TGAGGAGAGA AGCCATAGAA GATTTGACCA AGAAGAGAGA GGGCAATAAA
600
CCAAAGAGAC CCTTGAGATC ACGACATCCC GCAATTGTTT CTAGAGTAAT AGAAGGATTT
660
ACTCCGACAC TGCTACAATA AATTAAGGAA GACAAGGAAT TTGGTTTTTT TCATTGGAGG
720
AGTGTAATTT GTTTTTTGGC AAGCTCATCA CATGAATCAC ATGGAAAAAA AAAAAAA
777

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ATATGTTCAG AATTTCAAAT GTGGGAATGT CAACCTCCTT GAACTTCAGA ATTCAGGGCC
60
ATACGTTGAA GCTAGTCGAG GTTGAAGGAT CTCACACCGT CCAGAACATG TATGATTCAA
120
TCGATGTTCA CGTGGGCCAA TCCATGGCTG TCTTAGTGAC CTTAAATCAG CCTCCAAAGG
180
ACTACTACAT TGTCGCATCC ACCCGGTTCA CCAAGACGGT TCTCAATGCA ACTGCAGTGC
240
TACACTACAC CAACTCGCTT ACCCCAGTTT CCGGGCCACT ACCAGCTGGT CCAACTTACC
300
AAAAACATTG GTCCATGAAG CAAGCAAGAA CAATCAGGTG GAAC
344

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 341 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCCGCAACTG CAATTCTCTT CGTAAAACAT GACGGCTGTC GGCAAAACCT CTTTCCTCTT
60
GGGAGCTCTC CTCCTCTTCT CTGTGGCGGT GACATTGGCA GATGCAAAAG TTTACTACCA
120
TGATTTTGTC GTTCAAGCGA CCAAGGTGAA GAGGCTGTGC ACGACCCACA ACACCATCAC
180
GGTGAACGGG CAATTCCCGG GTCCGACTTT GGAAGTTAAC GACGGCGACA CCCTCGTTGT
240
CAATGTCGTC AACAAAGCTC GCTACAACGT CACCATTCAC TGGCACGGCG TCCGGCAGGT
300
GAGATCTGGT TGGGCTGATG GGGCGGAATT TGTGACTCAA T
341

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 358 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GAATTCGGCA CGAGATATGT TCAGAATTTC AAATGTGGGA ATGTCAACCT CCTTGAACTT 60
CAGAATTCAG GGCCATACGT TGAAGCTAGT CGAGGTTGAA GGATCTCACA CCGTCCAGAA 120
CATGTATGAT TCAATCGATG TTCACGTGGG CCAATCCATG GCTGTCTTAG TGACCTTAAA 180
TCAGCCTCCA AAGGACTACT ACATTGTCGC ATCCACCCGG TTCACCAAGA CGGTTCTCAA 240
TGCAACTGCA GTGCTACACT ACACCAACTC GCTTACCCCA GTTTCCGGGC CACTACCAGC 300
TGGTCCAACT TACCAAAAAC ATTGGTCCAT GAAGCAAGCA AGAACAATCA GGTGGAAC 358

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 409 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

ATCAAGAGTT TGAGTCTAAA CCTTGTCTAA TCCTCTCTCG CATAGTCATT TGGAGACGAA 60
TGCTGATCGG CCGCAGCTGC ATTCTCTTCG TAAAACATGA CGGCTGTCGG CAAAACCTCT 120
TTCCTCTTGG GAGCTCTCCT CCTCTTCTCT GTGGCGGTGA CATTGGCAGA TGCAAAAGTT 180
TACTACCATG ATTTTGTCGT TCAAGCGACC AAGGTGAAGA GGCTGTGCAC GACCCACAAC 240
ACCATCACGG TGAACGGGCA ATTCCCGGGT CCGACTTTGG AAGTTAACGA CGGCGACACC 300
CTCGTTGTCA ATGTCGTCAA CAAAGCTCGC TACAACGTCA CCATTCACTG GCACGGCGTC 360
CGGCAGGTGA GATCTGGTTG GGCTGATGGG GCGGAATTTG TGACTCAAT 409

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 515 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CTCTCTCTCT CTCTCTCTCT GTGTGTTCAT TCTCGTTGAG CTCGTGGTCG CCTCCCGCCA 60
TGGATCCGCA CAAGTACCGT CCATCCAGTG CTTTCAACAC TTCTTTCTGG ACTACGAACT 120

CTGGTGCTCC TGTCTGGAAC AATAACTCTT CGTTGACTGT TGGAAGCAGA GGTCCAATTC
180
TTCTTGAGGA TTATCACCTC GTGGAGAAAC TTGCCAACTT TGATAGGGAG AGGATTCCAG
240
AGCGTGTGGT GCATGCCAGA GGAGCCAGTG CAAAGGGATT CTTTGAGGTC ACTCATGACA
300
TTTCCCAGCT TACCTGTGCT GATTTCCTTC GGGCACCAGG AGTTCAAACA CCCGTGATTG
360
TCCGTTTCTC CACTGTCATC CACGAAAGGG GCAGCCCTGA AACCCTGAGG GACCCTCGAG
420
GTTTTGCTGT GAAGTTCTAC ACAAGAGAGG GTAACTTTGA TCTGGTGGGA AACAATTTCC
480
CTGTCTTCTT TGTCCGTAAT GGGATAAATT CCCCG
515

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 471 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GAATTCGGCA CGAGGCTCCC TCTCGTACTG CCATACTCCT GGGACGGGAT TCGGATAGGG
60
ATTTGCGGCG ATCCATTTCT CGATTCAAGG GGAAGAATCA TGGGGAAGTC CTACCCGACC
120
GTAAGCCAGG AGTACAAGAA GGCTGTCGAG AAATGCAAGA AGAAGTTGAG AGGCCTCATC
180
GCTGAGAAGA GCTGCGCTCC GCTCATGCTC CGCATCGCGT GGCACTCCGC CGGTACCTTC
240
GATGTGAAGA CGAAGACCGG AGGCCCGTTC GGGACCATGA AGCACGCCGC GGAGCTCAGC
300
CACGGGGCCA ACAGCGGGCT CGACGTTGCC GATCAGGTCT TGCAGCCGAT CAAGGATCAG
360
TTCCCCGTCA TCACTTATGC TGATTTCTAC CAGCTGGCTG GCGTCGTTGC TGTGGAAGTT
420
ACTGGTGGAC CTGAAGTTGC TTTTCACCCG GAAGAGAGGC AAACCACAAC C
471

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 487 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GAATTCGGCA CGAGCTCCCA CTTCTGTCTC GCCACCATTA CTAGCTTCAA AGCCCAGATC
60
TCAGTTTCGT GCTCTCTTCG TCATCTCTGC CTCTTGCCAT GGATCCGTAC AAGTATCGCC
120
CGTCCAGCGC TTACGATTCC AGCTTTTGGA CAACCAACTA CGGTGCTCCC GTCTGGAACA
180
ATGACTCATC GCTGACTGTT GGAACTAGAG GTCCGATTCT CCTGGAGGAC TACCATCTGA
240
TTGAGAAACT TGCCAACTTC GAGAGAGAGA GGATTCCTGA GCGGGTGGTC CATGCACGGG
300
GAGCCAGCGC GAAAGGGTTC TTCGAGGTCA CCCACGACAT CTCTCACTTG ACCTGTGCTG
360

ATTTCCTCCG GGCTCCTGGA GTCCAGACGC CCGTAATCGT CCGTTTCTCC ACCGTCATCC
420
ACGAGCGCGG CAGCCCGAAC CTCAGGGACC CTCGTGGTTT TGCAGTGAAG TTCTACACCA
480
GAGAGGG
487

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 684 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GAATTCCTGC AGCCCGGGGG ATCCACTAGT TCTAGAGCGG CCGCCACCGC GGTGGAGCTC
60
GCGCGCCTGC AGGTCGACAC TAGTGGATCC AAAGAATTCG GCACGAGGCC CGACGGCCAC
120
TTGTTGGACG CCATGGAAGC TCTCCGGAAA GCCGGGATTC TGGAACCGTT TAAACTGCAG
180
CCCAAGGAAG GACTGGCTCT CGTCAACGGC ACAGCGGTGG GATCCGCCGT GGCCGCGTCC
240
GTCTGTTTTG ACGCCAACGT GCTGGGCGTG CTGGCTGAGA TTCTGTCTGC GCTCTTCTGC
300
GAGGTGATGC AAGGGAAACC GGAGTTCGTA GATCCGTTAA CCCACCAGTT GAAGCACCAC
360
CCAGGGCAGA TCGAAGCCGC GGCCGTCATG GAGTTCCTCC TCGACGGTAG CGACTACGTG
420
AAAGAAGCAG CGCGGCTTCA CGAGAAAGAC CCGTTGAGCA AACCGAAACA AGACCGCTAC
480
GCTCTGCGAA CATCGCCACA GTGGTTGGGG CCTCCGATCG AAGTCATCCG CGCTGCTACT
540
CACTCCATCG AGCGGGAGAT CAATTCCGTC AACGACAATC CGTTAATCGA TGTCTCCAGG
600
GACATGGCTC TCCACGGCGG CAACTTCCAG GGAACACCCA TCGGAGTTTC CATGGACAAC
660
ATGCGAATCT CTTTGGCAGC CGTC
684

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 418 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GAATTCGGCA CGAGGACAAG GTCATAGGCC CTCTCTTCAA ATGCTTGGAT GGGTGGAAAG
60
GAACTCCTGG CCCATTCTGA AATAAATAAT CTTCCAAGAT CGCCTTTATA CAACGACTGC
120
TATGATTTGA GTCCTCGGAT CTTTTTGTTG ATGCAGTTGT TTACCGATCT GGAATTTGAT
180
TGGTCATAAA GCTTGATTTT GTTTTTCTTT CTTTTGTTTT ATACTGCTGG ATTTGCATCC
240
CATTGGATTT GCCAGAAATA TGTAAGGGTG GCAGATCATT TGGGTGATCT GAAACATGTA
300
AAAGTGGCGG ATCATTTGGG TAGCATGCAG ATCAGTTGGG TGATCGTGTA CTGCTTTCAC
360

TATTACTTAC ATATTTAAAG ATCGGGAATA AAAACATGAT TTTAATTGAA AAAAAAAA
418

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 479 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GATATCCCAA CGACCGAAAA CCTGTATTTT CAGGGCGCCA TGGGGATCCG GAATTCGGCA
60
CGAGCAAGGA AGAAAATATG GTTGCAGCAG CAGAAATTAC GCAGGCCAAT GAAGTTCAAG
120
TTAAAAGCAC TGGGCTGTGC ACGGACTTCG GCTCGTCTGG CAGCGATCCA CTGAACTGGG
180
TTCGAGCAGC CAAGGCCATG GAAGGAAGTC ACTTTGAAGA AGTGAAAGCG ATGGTGGATT
240
CGTATTTGGG AGCCAAGGAG ATTTCCATTG AAGGGAAATC TCTGACAATC TCAGACGTTG
300
CTGCCGTTGC TCGAAGATCG CAAGTGAAAG TGAAATTGGA TGCTGCGGCT GCCAAATCTA
360
GGGTCGAGGA GAGTTCAAAC TGGGTTCTCA CCCAGATGAC CAAGGGGACG GATACCTATG
420
GTGTCACTAC TGGTTTCGGA GCCACTTCTC ACAGGAGAAC GAACCAGGGA GCCGAGCTT
479

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1785 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

TATCGATAAG CTTGATATCG AATTCCTGCA GCCCGGGGGA TCCACTAGTT CTAGAGCGGC
60
CGCCACCGCG GTGGAGCTCG CGCGCCTGCA GGTCGACACT AGTGGATCCA AAGAATTCGG
120
CACGAGGTTG CAGGTCGGGG ATGATTTGAA TCACAGAAAC CTCAGCGATT TTGCCAAGAA
180
ATATGGCAAA ATCTTTCTGC TCAAGATGGG CCAGAGGAAT CTTGTGGTAG TTTCATCTCC
240
CGATCTCGCC AAGGAGGTCC TGCACACCCA GGGCGTCGAG TTTGGGTCTC GAACCCGGAA
300
CGTGGTGTTC GATATCTTCA CGGGCAAGGG GCAGGACATG GTGTTCACCG TCTATGGAGA
360
TCACTGGAGA AAGATGCGCA GGATCATGAC TGTGCCTTTC TTTACGAATA AAGTTGTCCA
420
GCACTACAGA TTCGCGTGGG AAGACGAGAT CAGCCGCGTG GTCGCGGATG TGAAATCCCG
480
CGCCGAGTCT TCCACCTCGG GCATTGTCAT CCGTAGGCGC CTCCAGCTCA TGATGTATAA
540
TATTATGTAT AGGATGATGT TCGACAGGAG ATTCGAATCC GAGGACGACC CGCTTTTCCT
600
CAAGCTCAAG GCCCTCAACG GAGAGCGAAG TCGATTGGCC CAGAGCTTTG AGTACAATTA
660
TGGGGATTTC ATTCCCATTC TTAGGCCCTT CCTCAGAGGT TATCTCAGAA TCTGCAATGA
720

GATTAAAGAG AAACGGCTCT CTCTTTTCAA GGACTACTTC GTGGAAGAGC GCAAGAAGCT
780
CAACAGTACC AAGACTAGTA CCAACACCGG GGGAGCTCAA GTGTGCAATG GACCATATTT
840
TAGATGCTCA GGACAAGGGA GAGATCAATG AGGATAATGT TTTGTACATC GTTGAGAACA
900
TCAACGTTGC AGCAATTGAG ACAACGCTGT GGTCGATGGA ATGGGGAATA GCGGAGCTGG
960
TGAACCACCA GGACATTCAG AGCAAGGTGC GCGCAGAGCT GGACGCTGTT CTTGGACCAG
1020
GCGTGCAGAT AACGGAACCA GACACGACAA GGTTGCCCTA CCTTCAGGCG GTTGTGAAGG
1080
AAACCCTTCG TCTCCGCATG GCGATCCCGT TGCTCGTCCC CCACATGAAT CTCCACGACG
1140
CCAAGCTCGG GGGCTACGAT ATTCCGGCAG AGAGCAAGAT CCTGGTGAAC GCCTGGTGGT
1200
TGGCCAACAA CCCCGCCAAC TGGAAGAACC CCGAGGAGTT CCGCCCCGAG CGGTTCTTCG
1260
AGGAGGAGAA GCACACCGAA GCCAATGGCA ACGACTTCAA ATTCCTGCCT TCGGTGTGGG
1320
GAGGAGGAGC TGCCCGGGAA TCATTCTGGC GCTGCCTCTC CTCGCACTCT CCATCGGAAG
1380
ACTTGTTCAG AACTTCCACC TTCTGCCGCC GCCCGGGCAG AGCAAAGTGG ATGTCACTGA
1440
GAAGGGCGGG CAGTTCAGCC TTCACATTCT CAACCATTCT CTCATCGTCG CCAAGCCCAT
1500
AGCTTCTGCT TAATCCCAAC TTGTCAGTGA CTGGTATATA AATGCGCGCA CCTGAACAAA
1560
AAACACTCCA TCTATCATGA CTGTGTGTGC GTGTCCACTG TCGAGTCTAC TAAGAGCTCA
1620
TAGCACTTCA AAAGTTTGCT AGGATTTCAA TAACAGACAC CGTCAATTAT GTCATGTTTC
1680
AATAAAAGTT TGCATAAATT AAATGATATT TCAATATACT ATTTTGACTC TCCACCAATT
1740
GGGGAATTTT ACTGCTAAAA AAAAAAAAAA AAAAAAAAAA AAAAA
1785

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 475 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GAATTCGGCA CGAGATTTCC ATGGACGATT CCGTTTGGCT TCAATTCGTT TCCTCTGGCT
60
GTCCTCGTCC TCGTTTTCCT TGTTCTTCCT CCGACTTTTT CTCTGGAAGC TATGGCGTAA
120
TAGGAACCTG CCGCCAGGAC CCCCGGCATG GCCGATCGTA GGGAACGTCC TTCAGATTGG
180
ATTTTCCAGC GGCGCGTTCG AGACCTCAGT GAAGAAATTC CATGAGAGAT ACGGTCCAAT
240
ATTCACTGTG TGGCTCGGTT CCCGCCCTCT GCTGATGATC ACCGACCGCG AGCTTGCCCA
300
CGAGGCGCTC GTACAGAAGG GCTCCGTCTT CGCTGACCGC CCGCCCGCCC TCGGGATGCA
360
GAAAATCTTC AGTAGCAACC AGCACAACAT CACTTCGGCT GAATACGGCC CGCTGTGGCG
420
GAGCCTTCGC AGGAATCTGG TTAAAGAAGC CCTGAGACTT CGGCGATGAA GGCTT
475

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 801 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GCTCCACCGA CGGTGGACGG TCCGCTACTC AGTAACTGAG TGGGATCCCC CGGGCTGACA
60
GGCAATTCGA TTTAGCTCAC TCATTAGGCA CCCCAGGCTT TACACTTTAT GCTTCCGGCT
120
CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTTCACA CAGGAAACAG CTATGACCAT
180
GATTACGCCA AGCGCGCAAT TAACCCTCAC TAAAGGGAAC AAAAGCTGGA GCTCCACCGC
240
GGTGGCGGCC GCTCTAGAAC TAGTGGATCC AAAGAATTCG GCACGAGACC CAGTGACCTT
300
CAGGCCTGAG AGATTTCTTG AGGAAGATGT TGATATTAAG GGCCATGATT ACAGGCTACT
360
GCCATTCGGT GCAGGGCGCA GGATCTGCCC TGGTGCACAA TTGGGTATTA ATTTAGTTCA
420
GTCTATGTTG GGACACCTGC TTCATCATTT CGTATGGGCA CCTCCTGAGG GAATGAAGGC
480
AGAAGACATA GATCTCACAG AGAATCCAGG GCTTGTTACT TTCATGGCCA AGCCTGTGCA
540
GGCCATTGCT ATTCCTCGAT TGCCTGATCA TCTCTACAAG CGACAGCCAC TCAATTGATC
600
AATTGATCTG ATAGTAAGTT TGAATTTTGT TTTGATACAA AACGAAATAA CGTGCAGTTT
660
CTCCTTTTCC ATAGTCAACA TGCAGCTTTC TTTCTCTGAA GCGCATGCAG CTTTCTTTCT
720
CTGAAGCCCA ACTTCTAGCA AGCAATAACT GTATATTTTA GAACAAATAC CTATTCCTCA
780
AATTGAGTAT TTCTCTGTAG G
801

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 744 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GGGCCCCCCT TCGAGGTGGA CACTAGTGGA TCCAAAGAAT TCGGCACGAG GTTTTATCTG
60
AAGGACGCTG TGCTTGAAGG CTCCCAGCCA TTCACCAAAG CCCATGGAAT GAATGCGTTC
120
GAGTACCCGG CCATCGATCA GAGATTCAAC AAGATTTTCA ACAGGGCTAT GTCTGAGAAT
180
TCTACCATGT TGATGAACAA GATTTTGGAT ACTTACGAGG GTTTTAAGGA GGTTCAGGAG
240
TTGGTGGATG TGGGAGGAGG TATTGGGTCG ACTCTCAATC TCATAGTGTC TAGGTATCCC
300
CACATTTCAG GAATCAACTT CGACTTGTCC CATGTGCTGG CCGATGCTCC TCACTACCCA
360
GCTGTGAAAC ATGTGGGTGG AGACATGTTT GATAGTGTAC CAAGTGGCCA AGCTATTTTT
420
ATGAAGTGGA TTCTGCATGA TTGGAGCGAT GATCATTGCA GGAAGCTTTT GAAGAATTGT
480

CACAAGGCGT TGCCAGAGAA GGGGAAGGTG ATTGCGGTGG ACACCATTCT CCCAGTGGCT
540
GCAGAGACAT CTCCTTATGC TCGTCAGGGA TTTCATACAG ATTTACTGAT GTTGGCATAC
600
AACCCAGGGG GCAAGGAACG CACAGAGCAA GAATTTCAAG ATTTAGCTAA GGAGACGGGA
660
TTTGCAGGTG GTGTTGAACC TGTATGTTGT GTCAATGGAA TGTGGGTAAT GGAATTCCTG
720
CAGCCCGGGG GATCCACTAG TTCT
744

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 426 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GTGGCCCTGG AAGTAGTGTG CGCGACATGG ATTCCTTGAA TTTGAACGAG TTTATGTTGT
60
GGTTTCTCTC TTGGCTTGCT CTCTACATTG GATTTCGTTA TGTTTTGAGA TCGAACTTGA
120
AGCTCAAGAA GAGGCGCCTC CCGCCGGGCC CATCGGGATG GCCAGTGGTG GGAAGTCTGC
180
CATTGCTGGG AGCGATGCCT CACGTTACTC TCTACAACAT GTATAAGAAA TATGGCCCCG
240
TTGTCTATCT CAAACTGGGG ACGTCCGACA TGGTTGTGGC CTCCACGCCC GCTGCAGCTA
300
AGGCGTTTCT GAAGACTTTG GATATAAACT TCTCCAACCG GCCGGGAAAT GCAGGAGCCA
360
CGTACATCGC CTACGATTCT CAGGACATGG TGTGGGCAGC GTATGGAGGA CGGTGGAAGA
420
TGGAGC
426

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 562 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CAGTTCGAAA TTAACCTCAC TAAAGGGAAC AAAAGCTGGA GTTCGCGCGC CTGCAGGTCG
60
ACACTAGTGG ATCCAAAGAA TTCGGCACGA GCTTTGAGGC AACCTACATT CATTGAATCC
120
CAGGATTTCT TCTTGTCCAA ACAGGTTTAA GGAAATGGCA GGCACAAGTG TTGCTGCAGC
180
AGAGGTGAAG GCTCAGACAA CCCAAGCAGA GGAGCCGGTT AAGGTTGTCC GCCATCAAGA
240
AGTGGGACAC AAAAGTCTTT TGCAGAGCGA TGCCCTCTAT CAGTATATAT TGGAAACGAG
300
CGTGTACCCT CGTGAGCCCG AGCCAATGAA GGAGCTCCGC GAAGTGACTG CCAAGCATCC
360
CTGGAACCTC ATGACTACTT CTGCCGATGA GGGTCAATTT CTGGGCCTCC TGCTGAAGCT
420
CATTAACGCC AAGAACACCA TGGAGATTGG GGTGTACACT GGTTACTCGC TTCTCAGCAC
480

AGCCCTTGCA TTGCCCGATG ATGGAAAGAT TCTAGCCATG GACATCAACA GAGAGAACTA
540
TGATATCGGA TTGCCTATAA TT
562

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1074 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

TCGTGCCGCT CGATCCTCAC AGGCCCTTTT TATTTCCCTG GTGAACGATA CGATGGGCTC
60
GCACGCTGAG AATGGCAACG GGGTGGAGGT TGTTGATCCA ACGGACTTAA CTGACATCGA
120
GAATGGGAAA CCAGGTTATG ACAAGCGTAC GCTGCCTGCG GACTGGAAGT TTGGAGTGAA
180
GCTTCAAAAC GTTATGGAAG AATCCATTTA CAAGTACATG CTGGAAACAT TCACCCGCCA
240
TCGAGAGGAC GAGGCGTCCA AGGAGCTCTG GGAACGAACA TGGAACCTGA CACAGAGAGG
300
GGAGATGATG ACATTGCCAG ATCAGGTGCA GTTCCTGCGC TTGATGGTAA AGATGTCAGG
360
TGCTAAAAAG GCATTGGAGA TCGGAGTTTT CACTGGCTAT TCATTGCTCA ATATCGCTCT
420
CGCTCTTCCT TCTGATGGCA AGGTGGTAGC TGTGGATCCA GGAGATGACC CCAAATTTGG
480
CTGGCCCTGC TTCGTTAAGG CTGGAGTTGC AGACAAAGTG GAGATCAAGA AAACTACAGG
540
GTTGGACTAT TTGGATTCCC TTATTCAAAA GGGGGAGAAG GATTGCTTCG ACTTTGCATT
600
CGTGGACGCA GACAAAGTGA ACTACGTGAA CTATCATCCA CGGCTGATGA AGTTAGTGCG
660
CGTGGGGGGC GTCATAATTT ACGACGACAC CCTCTGGTTT GGTCTGGTGG GAGGAAAGGA
720
TCCCCACAAC CTGCTTAAGA ATGATTACAT GAGGACTTCT CTGGAGGGTA TCAAGGCCAT
780
CAACTCCATG GTAGCCAACG ACCCCAACTT GGAGGTCGCC ACAGTCTTTA TGGGATATGG
840
TGTCACTGTT TGTTACCGCA CTGCTTAGTT AGCTAGTCCT CCGTCATTCT GCTATGTATG
900
TATATGATAA TGGCGTCGAT TTCTGATATA GGTGGTTTTT CAATGTTTCT ATCGTCATGT
960
TTTCTGTTTA GCCAGAATGT TTCGATCGTC ATGGTTTCTG TTAAAGCCAG AATAAAATTA
1020
GCCGCTTGCA GTTCAAAAAA AAAAAAAAAA AAAAACTCGA GACTAGTTCT CTTC
1074

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1075 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

TCGGAGCTCT CGAATCCTCA CAGGCCCTTT TTATTTCCCT GGTGAACGAT ACGATGGGCT
60

CGCACGCTGA GAATGGCAAC GGGGTGGAGG TTGTTGATCC AACGGACTTA ACTGACATCG
120
AAGAATGGGA AACCAGGTTA TGACAAGCGT CGCTGCCTGC GGACTGGAAG TTTGGAGTGA
180
AGCTTCAAAA CGTTATGGAA GAATCCATTT ACAAGTACAT GCTGGAAACA TTCACCCGCC
240
ATCGAGAGGA CGAGGCGTCC AAGGAGCTCT GGGAACGAAC ATGGAACCTG ACACAGAGAG
300
GGGAGATGAT GACATTGCCA GATCAGGTGC AGTTCCTGCG CTTGATGGTA AAGATGTCAG
360
GTGCTAAAAA GGCATTGGAG ATCGGAGTTT TCACTGGCTA TTCATTGCTC AATATCGCTC
420
TCGCTCTTCC TTCTGATGGC AAGGTGGTAG CTGTGGATCC AGGAGATGAC CCCAAATTTG
480
GCTGGCCCTG CTTCGTTAAG GCTGGAGTTG CAGACAAAGT GGAGATCAAG AAAACTACAG
540
GGTTGGACTA TTTGGATTCC CTTATTCAAA AGGGGGAGAA GGATTGCTTC GACTTTGCAT
600
TCGTGGACGC AGACAAAGTG AACTACGTGA ACTATCATCC ACGGCTGATG AAGTTAGTGC
660
GCGTGGGGGG CGTCATAATT TACGACGACA CCCTCTGGTT TGGTCTGGTG GGAGGAAAGG
720
ATCCCCACAA CCTGCTTAAG AATGATTACA TGAGGACTTC TCTGGAGGGT ATCAAGGCCA
780
TCAACTCCAT GGTAGCCAAC GACCCCAACT TGGAGGTCGC CACAGTCTTT ATGGGATATG
840
GTGTCACTGT TTGTTACCGC ACTGCTTAGT TAGCTAGTCC TCCGTCATTC TGCTATGTAT
900
GTATATGATA ATGGCGTCGA TTTCTGATAT AGGTGGTTTT TCAATGTTTC TATCGTCATG
960
TTTTCTGTTT AGCCAGAATG TTTCGATCGT CATGGTTTCT GTTAAAGCCA GAATAAAATT
1020
AGCCGCTTGC AGTTCAAAAA AAAAAAAAAA AAAAAACTCG AGACTAGTTC TCTTC
1075

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1961 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GTTTTCCGCC ATTTTTCGCC TGTTTCTGCG GAGAATTTGA TCAGGTTCGG ATTGGGATTG
60
AATCAATTGA AAGGTTTTTA TTTTCAGTAT TTCGATCGCC ATGGCCAACG GAATCAAGAA
120
GGTCGAGCAT CTGTACAGAT CGAAGCTTCC CGATATCGAG ATCTCCGACC ATCTGCCTCT
180
TCATTCGTAT TGCTTTGAGA GAGTAGCGGA ATTCGCAGAC AGACCCTGTC TGATCGATGG
240
GGCGACAGAC AGAACTTATT GCTTTTCAGA GGTGGAACTG ATTTCTCGCA AGGTCGCTGC
300
CGGTCTGGCG AAGCTCGGGT TGCAGCAGGG GCAGGTTGTC ATGCTTCTCC TTCCGAATTG
360
CATCGAATTT GCGTTTGTGT TCATGGGGGC CTCTGTCCGG GGCGCCATTG TGACCACGGC
420
CAATCCTTTC TACAAGCCGG GCGAGATCGC CAAACAGGCC AAGGCCGCGG GCGCGCGCGA
480
TCATAGTTAC CCTGGCAGCT TATGTGGAGA AACTGGCCGA TCTGCAGAGC CACGATGTGC
540
TCGTCATCAC AATCGATGAT GCTCCCAAGG AAGGTTGCCA ACATATTTCC GTTCTGACCG
600

AAGCCGACGA AACCCAATGC CCGGCCGTGA CAATCCACCC GGACGATGTC GTGGCGTTGC
660
CCTATTCTTC CGGAACCACG GGGCTCCCCA AGGGCGTGAT GTTAACGCAC AAAGGCCTGG
720
TGTCCAGCGT TGCCCAGCAG GTCGATGGTG AAAATCCCAA TCTGTATTTC CATTCCGATG
780
ACGTGATACT CTGTGTCTTG CCTCTTTTCC ACATCTATTC TCTCAATTCG GTTCTCCTCT
840
GCGCGCTCAG AGCCGGGGCT GCGACCCTGA TTATGCAGAA ATTCAACCTC ACGACCTGTC
900
TGGAGCTGAT TCAGAAATAC AAGGTTACCG TTGCCCCAAT TGTGCCTCCA ATTGTCCTGG
960
ACATCACAAA GAGCCCCATC GTTTCCCAGT ACGATGTCTC GGCCGTCCGG ATAATCATGT
1020
CCGGCGCTGC GCCTCTCGGG AAGGAACTCG AAGATGCCCT CAGAGAGCGT TTTCCCAAGG
1080
CCATTTTCGG GCAGGGCTAC GGCATGACAG AAGCAGGCCC GGTGCTGGCA ATGAACCTAG
1140
CCTTCGCAAA GAATCCTTTC CCCGTCAAAT CTGGCTCCTG CGGAACAGTC GTCCGGAACG
1200
CTCAAATAAA GATCCTCGAT ACAGAAACTG GCGAGTCTCT CCCGCACAAT CAAGCCGGCG
1260
AAATCTGCAT CCGCGGACCC GAAATAATGA AAGGATATAT TAACGACCCG GAATCCACGG
1320
CCGCTACAAT CGATGAAGAA GGCTGGCTCC ACACAGGCGA CGTCGGGTAC ATTGACGATG
1380
ACGAAGAAAT CTTCATAGTC GACAGAGTAA AGGAGATTAT CAAATATAAG GGCTTCCAGG
1440
TGGCTCCTGC TGAGCTGGAA GCTTTACTTG TTGCTCATCC GTCAATCGCT GACGCAGCAG
1500
TCGTTCCTCA AAAGCACGAG GAGGCGGGCG AGGTTCCGGT GGCGTTCGTG GTGAAGTCGT
1560
CGGAAATCAG CGAGCAGGAA ATCAAGGAAT TCGTGGCAAA GCAGGTGATT TTCTACAAGA
1620
AAATACACAG AGTTTACTTT GTGGATGCGA TTCCTAAGTC GCCGTCCGGC AAGATTCTGA
1680
GAAAGGATTT GAGAAGCAGA CTGGCAGCAA AATGAAAATG AATTTCCATA TGATTCTAAG
1740
ATTCCTTTGC CGATAATTAT AGGATTCCTT TCTGTTCACT TCTATTTATA TAATAAAGTG
1800
GTGCAGAGTA AGCGCCCTAT AAGGAGAGAG AGAGCTTATC AATTGTATCA TATGGATTGT
1860
CAACGCCCTA CACTCTTGCG ATCGCTTTCA ATATGCATAT TACTATAAAC GATATATGTT
1920
TTTTTTATAA ATTTACTGCA CTTCTCGTTC AAAAAAAAAA A
1961

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1010 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GACAAACTTG GTCGTTTGTT TAGGTTTTGC TGCAGGTGAA CACTAATATG GAAGGCCAGA
60
TTGCAGCATT AAGCAAAGAA GATGAGTTCA TTTTTCACAG CCCTTTTCCT GCAGTACCTG
120
TTCCAGAGAA TATAAGTCTT TTCCAGTTTG TTCTGGAAGG TGCTGAGAAA TACCGTGATA
180
AGGTGGCCCT CGTGGAGGCC TCCACAGGGA AGGAGTACAA CTATGGTCAG GTGATTTCGC
240

TCACAAGGAA TGTTGCAGCT GGGCTCGTGG ACAAAGGCAT TCAAAAGGGC GATGTTGTAT
300
TTGTTCTGCT TCCAAATATG GCAGAATACC CCATTATTGT GCTGGGAATA ATGTTGGCCG
360
GCGCAGTGTT TTCTGGGGCA AATCCTTCTG CACACATCAA TGAAGTTGAA AAACATATCC
420
AGGATTCTGG AGCAAAGATT GTTGTGACAG TTGGGTCTGC TTATGAGAAG GTGAGGCAAG
480
TGAAACTGCC TGTTATTATT GCAGATAACG AGCATGTCAT GAACACAATT CCATTGCAGG
540
AAATTTTTGA GAGAAACTAT GAGGCCGCAG GGCCTTTTGT ACAAATTTGT CAGGATGATC
600
TGTGTGCACT CCCTTATTCC TCTGGCACCA CAGGGGCCTC TAAAGGTGTC ATGCTCACTC
660
ACAGAAATCT GATTGCAAAT CTGTGCTCTA GCTTGTTTGA TGTCCATGAA TCTCTTGTAG
720
GAAATTTCAC CACGTTGGGG CTGATGCCAT TCTTTCACAT ATATGGCATC ACGGGCATCT
780
GTTGCGCCAC TCTTCGCAAC GGAGGCAAGG TCGTGGTCAT GTCCAGATTC GATCTCCGAC
840
ACTTTATCAG TTCTTTGATT ACTTATGAGG TCAACTTCGC GCCTATTGTC CCGCCTATAA
900
TGCTCTCCCT CCGGTTTAAA AATCCTATCG TTAACGAGTT CGATCTCAGC CGCTTGAAAC
960
TCCAAAGCTG TTCATGACTG CGGCTGCTCC ACTGGCGCCG GATCTACTGC
1010

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 741 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GAATTCGGCA CGAGACCATT TCCAGCTAAT ATTGGCATAG CAATTGGTCA TTCTATCTTT
60
GTCAAAGGAG ATCAAACAAA TTTTGAAATT GGACCTAATG GTGTGGAGGC TAGTCAGCTA
120
TACCCAGATG TGAAATATAC CACTGTCGAT GAGTACCTCA GCAAATTTGT GTGAAGTATG
180
CGAGATTCTC TTCCACATGC TTCAGAGATA CATAACAGTT TCAATCAATG TTTGTCCTAG
240
GCATTTGCCA AATTGTGGGT TATAATCCTT CGTAGGTGTT TGGCAGAACA GAACCTCCTG
300
TTTAGTATAG TATGACGAGC TAGGCACTGC AGATCCTTCA CACTTTTCTC TTCCATAAGA
360
AACAAATACT CACCTGTGGT TTGTTTTCTT TCTTTCTGGA ACTTTGGTAT GGCAATAATG
420
TCTTTGGAAA CCGCTTAGTG TGGAATGCTA AGTACTAGTG TCCAGAGTTC TAAGGGAGTT
480
CCAAAATCAT GGCTGATGTG AACTGGTTGT TCCAGAGGGT GTTTACAACC AACAGTTGTT
540
CAGTGAATAA TTTTGTTAGA GTGTTTAGAT CCATCTTTAC AAGGCTATTG AGTAAGGTTG
600
GTGTTAGTGA ACGGAATGAT GTCAAATCTT GATGGGCTGA CTGACTCTCT TGTGATGTCA
660
AATCTTGATG GATTGTGTCT TTTTCAATGG TAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
720
AAAAAAAAAA AAAAAAAAAA A
741

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 643 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CTCATCTCGG AGTTGCAGGC TGCAGCTTTT GGCCCAAAGC ATGATATCAG ATCAAACGAC
60
GCAGATGAAG CAAACGGATC AAACAGTTTG CGTTACTGGA GCAGCGGGTT TCATTGCCTC
120
ATGGCTTGTC AAGATGCTCC TCATCAGAGG TTACACTGTC AGAGCAGCAG TTCGGACCAA
180
CCCAGCTGAT GATAGGTGGA AGTATGAGCA TCTGCGAGAG TTGGAAGGAG CAAAAGAGAG
240
GCTTGAGCTT GTGAAAGCTG ATATTCTCCA TTACCAGAGC TTACTCACAG TCATCAGAGG
300
TTGCCACGGT GTCTTTCACA TGGCTTCAGT TCTCAATGAT GACCCTGAGC AAGTGATAGA
360
ACCAGCAGTC GAAGGGACGA GGAATGTGAT GGAGGCCTGC GCAGAAACTG GGGTGAAGCG
420
CGTTGTTTTT ACTTCTTCCA TCGGCGCAGT TTACATGAAT CCTCATAGAG ACCCGCTCGC
480
GATTGTCCAT GATGACTGCT GGAGCGATTT GACTACTGCG TACAAACCAA GAATTGGTAT
540
TGCTATGCAA AAACCTTGGC AGAGAAATCT GCATGGGATA TTGCTAAGGG AAGGAATTTA
600
GAGCTTGCAG TGATAAATCC AGGCCTGGCC TTAGGTCCCT TGA
643

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GAATTCGGCA CGAGAATTTT TCTGTGGTAA GCATATCTAT GGCTCAAACC AGAGAGAAGG
60
ACGATGTCAG CATAACAAAC TCCAAAGGAT TGGTATGCGT GACAGGAGCG GCTGGTTACT
120
TGGCATCTTG GCTTATCAAG CGTCTCCTCC AGTGTGGTTA CCAAGTGAGA GGAACTGTGC
180
GGGATCCTGG CAATGAGAAA AAGATGGCTC ATTTATGGAA GTTAGATGGG GCGAAAGAGA
240
GACTGCAACT AATGAAAGCT GATTTAATGG ACGAGGGCAG CTTCGATGAG GTCATCAGAG
300
GCTGCCATGG TGTTTTTCAC ACAGCGTCTC CAGTCGTGGG TGTCAAATCA GATCCCAAGA
360
TATGGTATGC TCTGGCCAAG ACTTTAGCAG AAAAAGCAGC ATGGGATTTT GCCCAAGAAA
420
ACCATCTGGA CATGGTTGCA G
441

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 913 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GAATTCGGCA CGAGGAAAAC ATCATCCAGG CATTTTGGAA ATTTAGCTCG CCGGTTGATT
60
CAGGATCCTG CAATGGCTTT TGGCGAAGAG CAGACTGCCT TGCCACAAGA AACGCCTTTG
120
AATCCTCCGG TCCATCGAGG AACAGTGTGC GTTACAGGAG CTGCTGGGTT CATAGGGTCA
180
TGGCTCATCA TGCGATTGCT TGAGCGAGGA TATAGTGTTA GAGCAACTGT GCGAGACACT
240
GGTAATCCTG TAAAGACAAA GCATCTGTTG GATCTGCCGG GGGCAAATGA GAGATTGACT
300
CTCTGGAAAG CAGATTTGGA TGATGAAGGA AGCTTTGATG CTGCCATTGA TGGGTGTGAG
360
GGTGTTTTCC ATGTTGCCAC TCCCATGGAT TTCGAGTCCG AGGATCCCGA GAATGAGATA
420
ATTAAGCCAA CAATCAACGG GGTCTTGAAT GTTATGAGAT CGTGTGCAAA AGCCAAGTCC
480
GTGAAGCGAG TTGTTTTCAC GTCATCTGCT GGGACTGTGA ATTTTACAGA TGATTTCCAA
540
ACACCAGGCA AAGTTTTTGA CGAATCATGC TGGACCAACG TGGATCTTTG CAGAAAAGTT
600
AAAATGACAG GATGGATGTA CTTTGTATCG AAGACATTAG CAGAGAAAGC TGCTTGGGAT
660
TTTGCAGAGG AGAACAAGAT CGATCTCATT ACTGTTATCC CCACATTGGT CGTTGGACCA
720
TTCATTATGC AGACCATGCC ACCGAGCATG ATCACAGCCT TGGCACTGTT AACGCGGAAT
780
GAACCCCACT ACATGATACT GAGACAGGTA CAGCTGGTTC ACTTGGATGA TCTCTGTATG
840
TCACATATCT TTGTATATGA ACATCCTGAA GCAAAGGGCA GATACATCTC TTCCACATGT
900
GATGCTACCC ATT
913

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 680 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GAATTCGGCA CGAGATCAAT TTTTGCATAT TATTAAAAAG TAAGTGTATT CGTTCTCTAT
60
ATTGATCAGT CACAGAGTCA TGGCCAGTTG TGGTTCCGAG AAAGTAAGAG GGTTGAATGG
120
AGATGAAGCA TGCGAAGAGA ACAAGAGAGT GGTTTGTGTA ACTGGGGCAA ATGGGTACAT
180
CGGCTCTTGG CTGGTCATGA GATTACTGGA ACATGGCTAT TATGTTCATG GAACTGTTAG
240
GGACCCAGAA GACACAGGGA AGGTTGGGCA TTTGCTGCGG CTCCCAGGGG CAAGTGAGAA
300
GCTAAAGCTG TTCAAGGCAG AGCTTAACGA CGAAATGGCC TTTGATGATG CTGTGAGCGG
360
TTGTCAAGGG GTTTTCCACG TTGCCAAGCC TGTTAATCTG GACTCAAACG CTCTTCAGGG
420
GGAGGTTGTT GGTCCTGCGG TGAGGGGAAC AGTAAATCTG CTTCGAGCCT GCGAACGATC
480

GGGCACTGTG AAACGAGTGA TACATACCTC GTCCGTTTCA GCAGTGAGAT TCACTGGGAA
540
ACCTGACCCC CCTGATACTG TGCTGGATGA ATCTCATTGG ACTTCGGTCG AGTATTGCAG
600
AAAGACAAAG ATGGTCGGAT GGATGTACTA CATCGCCAAC ACTTATGCAG AAGAGGGAGC
660
CCATAAGTTC GGATCAGAGA
680

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 492 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GAATTCGGCA CGAGGCTGGT TCAAGTGTCA GCCCAATGGC CTCCCCTACA GAGAATCCCC
60
AGATTTCAGA AGAGCTGCTA AATCATGAGA TCCATCAAGG AAGTACAGTA TGTGTGACAG
120
GAGCTGCTGG CTTCATAGGA TCATGGCTCG TCATGCGTTT GCTTGAGCGA GGATATACTG
180
TTAGAGGAAC TGTGCGAGAC ACTGGTAATC CGGTGAAGAC GAAGCATCTA TTGGATCTGC
240
CTGGGGCGAA TGAGAGGTTA ACTCTCTGGA AAGCAGATTT GGATGATGAA GGAAGCTTTG
300
ACGCCGCCAT TGATGGTTGT GAGGGAGTTT TCCATGTTGC CACTCCCATG GATTTTGAAT
360
CCGAGGACCC CGAGAACGAG ATAATTAAAC CCGCTGTCAA TGGGATGTTG AATGTTTTGA
420
GATCGTGTGG GAAAACCAAG TCTATGAAGC GAGTTGTTTT CACGTCGTCT GCTGGGACTC
480
TGCTTTTTAC GG
492

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 524 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GAATTCGGCA CGAGCTTGTT CAAAGTCACA TATCTTATTT TCTTTGTGAT ATCTGCAATT
60
TCCAAGCTTT TCGTCTACCT CCCTGAAAAG ATGAGCGAGG TATGCGTGAC AGGAGGCACA
120
GGCTTCATAG CTGCTTATCT CATTCGTAGT CTTCTCCAGA AAGGTTACAG AGTTCGCACT
180
ACAGTTCGCA ACCCAGATAA TGTGGAGAAG TTTAGTTATC TGTGGGATCT GCCTGGTGCA
240
AACGAAAGAC TCAACATCGT GAGAGCAGAT TTGCTAGAGG AAGGCAGTTT TGATGCAGCA
300
GTAGATGGTG TAGATGGAGT ATTCCATACT GCATCACCTG TCTTAGTCCC ATATAACGAG
360
CGCTTGAAGG AAACCCTAAT AGATCCTTGT GTGAAGGGCA CTATCAATGT CCTCAGGTCC
420
TGTTCAAGAT CACCTTCAGT AAAGCGGGTG GTGCTTACAT CCTCCTGCTC ATCAATACCG
480

ATACGACTAT AATAGCTTAG AGCGTTCCCT GCTGGACTGA GTCA
524

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 417 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

TCCTAATTGT TCGATCCTCC CTTTTAAAGC CCTTCCCTGG CCTTCATTCC AGGTCACAGA
60
GTTGTTCATG CAGTGCTAGC AGGAGGAGCA GCGTTGCAAT TGGGGAAAAT TCCAAAATCA
120
ATAACGAGAG GACAGAAGTA AGTTTGTGGA AATAGCAACC ATGCCGGTGT TTCCTTCTGG
180
TCTGGACCCC TCTGAGGACA ATGGCAAGCT CGTTTGTGTC ATGGATGCGT CCAGTTATGT
240
AGGTTTGTGG ATTGTTCAGG GCCTTCTTCA ACGAGGCTAT TCAGTGCATG CCACGGTGCA
300
GAGAGACGCT GGCGAGGTTG AGTCTCTCAG AAAATTGCAT GGGGATCGAT TGCAGATCTT
360
CTATGCAGAT GTCTTGGATT ATCACAGCAT TACTGATGCG CTCAAGGGCT GTTCTGG
417

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 511 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

ATGACACGAA TTTGTGCCTC TCTCTGACCA GAGCTTGAAG CTCTGTCTTC TCTGATATCG
60
CTTCATTCCA TCATCCAGGA GCTTCTGTTA TATCCATTTC CTCAAAATGG ATGCCTACCT
120
TGAAGAAAAT GGATACGGCG CTTCCAATTC TCGGAAATTA ATGTGCCTTA CCGGGGGCTG
180
GAGTTTCCTG GGGATTCATA TCGCAAGAAT GCTGCTCGGC CGGGGTTACT CAGTCCGTTT
240
CGCAATTCCG GTAACGCCAG AAGAGGCAGG CTCACTTATG GAATCCGAAG AAGCATTATC
300
GGGGAAGCTG GAGATATGCC AAGCCGATCT CTTGGATTAT CGCAGCGTTT TCGGCAACAT
360
CAATGGTTGC TCCGGAGTCT TCCACGTCCC TGCGCCCTGT GATCATCTGG ATGGATTACA
420
GGAGTATCCG GTATGATTAG TTTAATAGAT TGACGGGGTA TCCTGTATGA ATTAGTTTAT
480
GAATTTAAGG TTTTCTTAGA ATTTGGATAC T
511

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 609 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CATTGATAGT TGATGGAAGA CCATCAGTAA AGCATGAAAA AGAAATTGTT CCAAGGTGAA
60
GAAGTCAGTT GCTCCAGCAG AACCTTTTTA GCAATTGTTT TTGTATCCTT TTTGCCTTTG
120
AATATGTAAT CCATAAACTT ATGCAGGAAG TGCCTCGTGC CGAATTCGGC ACGAGAATCA
180
CTGACCTTCA CATATTTATT CCAATTCTAA TATCTCTACT CGCTGTCTAC CTGATTTTTC
240
AGTGGCGAAC CAACTTGACA GGGTTGGACA TGGCCAACAG CAGCAAGATT CTGATTATTG
300
GAGGAACAGG CTACATTGGT CGTCATATAA CCAAAGCCAG CCTTGCTCTT GGTCATCCCA
360
CATTCCTTCT TGTCAGAGAG ACCTCCGCTT CTAATCCTGA GAAGGCTAAG CTTCTGGAAT
420
CCTTCAAGGC CTCAGGTGCT ATTATACTCC ATGGATCTTT GGAGGACCAT GCAAGTCTTG
480
TGGAGGCAAT CAAGAAAGTT GATGTAGTTA TCTCGGCTGT CAAGGGACCA CAGCTGACGG
540
TTCAAACAGG ATATTTATCC AGGGTATTTA AAGGGAGGGT TGGAACCCAT CAAGAAGGGT
600
TTTGGCCAA
609

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GCAAGATAGG TTTTATTCTT CTGGAGTTGG GTGAGGCTTG GAAATTTAAG TAAAAAGGGT
60
GCATAGCAAT TAAGCAGTTG CAGCCATGGC GGTCTGTGGA ACTGAAGTAG CTCATACTGT
120
GCTCTATGTA GCTGCAGACA TGGTGGAAAA CAACACGTCT ATTGTGACCA CCTCTATGGC
180
TGCAGCAAAT TGTGAGATGG AGAAGCCTCT TCTAAATTCC TCTGCCACCT CAAGAATACT
240
GGTGATGGGA GCCACAGGTT ACATTGGCCG TTTTGTTGCC CAAGAAGCTG TTGCTGCTGG
300
TCATCCTACC TATGCTCTTA TACGCCCGTT TGCTGCTTGT GACCTGGCCA AAGCACAGCG
360
CGTCCAACAA TTGAAGGATG CCGGGGTCCA TATCCTTTAT GGGTCTTTGA GTGATCACAA
420
CCTCTTAGTA AATACATTGA AGGACATGGG CCGTTGTTAT CTCTACCATT GGAG
474

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

GCAAGATAGG TTTTATTCTT CTGGAGTTGG GTGAGGCTTG GAAATTTAAG TAAAAAGGGT
60
GCATAGCAAT TAAGCAGTTG CAGCCATGGC GGTCTGTGGA ACTGAAGTAG CTCATACTGT
120
GCTCTATGTA GCTGCAGACA TGGTGGAAAA CAACACGTCT ATTGTGACCA CCTCTATGGC
180
TGCAGCAAAT TGTGAGATGG AGAAGCCTCT TCTAAATTCC TCTGCCACCT CAAGAATACT
240
GGTGATGGGA GCCACAGGTT ACATTGGCCG TTTTGTTGCC CAAGAAGCTG TTGCTGCTGG
300
TCATCCTACC TATGCTCTTA TACGCCCGTT TGCTGCTTGT GACCTGGCCA AAGCACAGCG
360
CGTCCAACAA TTGAAGGATG CCGGGGTCCA TATCCTTTAT GGGTCTTTGA GTGATCACAA
420
CCTCTTAGTA AATACATTGA AGGACATGGG CCGTTGTTAT CTCTACCATT GGAG
474

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 608 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CATTGATAGT TGATGGAAGA CCATCAGTAA AGCATGAAAA AGAAATTGTT CCAAGGTGAA
60
GAAGTCAGTT GCTCCAGCAG AACCTTTTTA GCAATTGTTT TTGTATCCTT TTTGCCTTTG
120
AATATGTAAT CCATAAACTT ATGCAGGAAG TGCCTCGTGC CGAATTCGGC ACGAGAATCA
180
CTGACCTTCA AATATTTATT CCAATTCTAA TATCTCTACT CGCTGTCTAC CTGATTTTTC
240
AGTGGCGAAC CAACTTGACA GGGTTGGACA TGGCCAACAG CAGCAAGATT CTGATTATTG
300
GAGGAACAGG CTACATTGGT CGTCATATAA CCAAAGCCAG CCTTGCTCTT GGTCATCCCA
360
CATTCCTTCT TGTCAGAGAG ACCTCCGCTT CTAATCCTGA GAAGGCTAAG CTTCTGGAAT
420
CCTTCAAGGC CTCAGGTGCT ATTATACTCC ATGGATCTTT GGAGGACCAT GCAAGTCTTG
480
TGGAGGCAAT CAAGAAAGTT GATGTAGTTA TCTCGGCTGT CAAGGGACCA CAGCTGACGG
540
ATCAAACAGG ATATTTATCC AGGGTATTTA AAGGGAGGTT GGAACCCATC AAGAAGGGTT
600
TTGGCCAA
608

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GAATTCGGCA CGAGAAAACG TCCATAGCTT CCTTGCCAAC TGCAAGCAAT ACAGTACAAG
60
AGCCAGACGA TCGAATCCTG TGAAGTGGTT CTGAAGTGAT GCGAAGCTTG GAATCTGAAA
120

AAACTGTTAC AGGATATGCA GCTCGGGACT CCAGTGGCCA CTTGTCCCCT TACACTTACA
180
ATCTCAGAAA GAAAGGACCT GAGGATGTAA TTGTAAAGGT CATTTACTGC GGAATCTGCC
240
ACTCTGATTT AGTTCAAATG CGTAATGAAA TGGACATGTC TCATTACCCA ATGGTCCCTG
300
GGCATGAAGT GGTGGGGATT GTAACAGAGA TTGGCAGCGA GGTGAAGAAA TTCAAAGTGG
360
GAGAGCATGT AGGGGTTGGT TGCATTGTTG GGTCCTGTCG CAGTTGCGGT AATTGCAATC
420
AGAGCATGGA ACAATACTGC AGCAAGAGGA TTTGGACCTA CAATGATGTG AACCATGACG
480
GCACACCTAC TCAGGGCGGA TTTGCAAGCA GTATGGTGGT TGATCAGATG TTTGTGGTTC
540
GAATCCCGGA GAATCTTCCT CTGGAACAAG CGGCCCCTCT GTTATGTGCA GGGGTTACAG
600
TTTTCAGCCC AATGAAGCAT TTCGCCATGA CAGAGCCCGG GAAGAAATGT GGGATTTTGG
660
GTTTAGGAGG CGTGGGGCAC ATGGGTGTCA AGATTGCCAA AGCCTTTGGA CTCCACGTGA
720
CGGTTATCAG TTCGTCTGAT AAAAAGAAAG AAGAAGCCAT GGAAGTCCTC GGCGCCGATG
780
CTTATCTTGT TAGCAAGGAT ACTGAAAAGA TGATGGAAGC AGCAGAGAGC CTAGATTACA
840
TAATGGACAC CATTCCAGTT GCTCATCCTC TGGAACCATA TCTTGCCCTT CTGAAGACAA
900
ATGGAAAGCT AGTGATGCTG GGCGTTGTTC CAGAGCCGTT GCACTTCGTG ACTCCTCTCT
960
TAATACTTGG GAGAAGGAGC ATAGCTGGAA GTTTCATTGG CAGCATGGAG GAAACACAGG
1020
AAACTCTAGA TTTCTGTGCA GAGAAGAAGG TATCATCGAT GATTGAGGTT GTGGGCCTGG
1080
ACTACATCAA CACGGCCATG GAAAGGTTGG AGAAGAACGA TGTCCGTTAC AGATTTGTGG
1140
TGGATGTTGC TAGAAGCAAG TTGGATAATT AGTCTGCAAT CAATCAATCA GATCAATGCC
1200
TGCATGCAAG ATGAATAGAT CTGGACTAGT AGCTTAACAT GAAAGGGAAA TTAAATTTTT
1260
ATTTAGGAAC TCGATACTGG TTTTTGTTAC TTTAGTTTAG CTTTTGTGAG GTTGAAACAA
1320
TTCAGATGTT TTTTTAACTT GTATATGTAA AGATCAATTT CTCGTGACAG TAAATAATAA
1380
TCCAATGTCT TCTGCCAAAT TAATATATGT ATTCGTATTT TTATATGAAA AAAAAAAAAA
1440
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA
1474

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1038 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GAATTCGGCA CGAGAGAGGG TTATATATCT TGATTCTGAC CTGATTGTCG TCGACGACAT
60
TGCCAAGCTC TGGGCCACGG ATTTGGAATC TCGTGTCCTC GGGGCACCAG AGTACTGCAA
120
GGCGAATTTC ACAAAGTATT TCACCGATAA TTTCTGGTGG GATCCCGCAT TATCCAAGAC
180
CTTTGAGGGA AAAAAACCCT GCTACTTCAA CACAGGCGTA ATGGTGATCG ATCTTGAAAA
240

ATGGCGGGCA GGGGAATTCA CAAGAAAGAT CGAAATCTGG ATGGACATAC AGAAGGAACG
300
CCGTATCTAT GAGCTCGGAT CATTACCGCC ATTTTTACTG GTATTTGCTG GTTTGGTTAA
360
GCAAGTCGAT CATCGTTGGA ATCAGCACGG TTTAGGCGGA GATAATTTGC AAGGCCTTTG
420
CCGAGATCTT CACCCTGGAC CTGTCAGTTT GTTGCATTGG AGTGGTAAGG GCAAACCTTG
480
GCTACGCCTG GAATGCCAAG CGGACTTGCC CTCTGGATAC TTTATGGGCT CCTTATGATC
540
TTTATCGATC AACGTATTAC CTAAATGGGT GAGAGAGCCT CTCTCCTCGG GGTGCTTTTT
600
ATCGAATTAA ACCTGATTTG ATAAAATGCC AAATAGAACT TTACGCCTAT GCATCTTTCA
660
GTTTTGAATT TCAATTCTGG TAACGAATAG AAGAAAACAA TAGCACAGCC ACAGGCAGGA
720
CAAATCCATC ATGAGGGACC AATCGTTTGA ATTTAGTATT AATAAGGTTG TTCCATATAA
780
CGCCTGTGAA GAATGATATT GTGGACTGAT CTATTTATAT TTGTACTGCC ATGCCATCCT
840
CAGCCAGCAG AGAGGCAAGC AATGCCGCTG CAAGTCATGT AGGGAAGGCG TTGTGAACTC
900
AATTTTCGGC GACTGTACAG GATGTAAATT TTTGGAACAT TAATATCATT ATGATAAGTT
960
CCTGAACCAA CAACTGTATA ATACCTTATA AATGTATCTG CAACTCCATT TTTGCATAAA
1020
AAAAAAAAAA AAAAAAAA
1038

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CTAGGGGTCT TGGGGGGTTC CTGATGCCCA ATTGTTGCTG TGCTTGGCAT GAACCCAAAA
60
CATGCAAGAG ATCTGTAGTC AGTAGTCTTG TTGGATCTAT AGCTTTTAGA AAAGAGTCAC
120
GTCCTTTTAG GGTAACATCA TTCCAACCAT ATCCAGTTCC ACCACCGGCT ACACCTTCAA
180
CGGGAGGAGG AGCAAGATAT TCAGCATTGC TTTGGGCACC AGATGGATAG GCATTATTTT
240
CCATCGGAAT TCAGCCGAGC TCGCCCCCTC AGTCCAATCG TCGTGAAAAT CCCTCAAAAT
300
TGGGCAATTC TGGCTCGAAA TCGCCAAATT ATGGGCTACA ACAGGATTAA AATTGCACAG
360
AAATCTGCCA GT
372

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 545 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

AAAGAATTCG GCACGAGGGC AATCCGAGCC TAGCCAACCA ACTTGGCAGC AAGGAGCACA
60
GGGAGTTGGC GAGAGAAGCT GTTAGGAAAT CTTTGGTATT GTTGAAAAAT GGGAAGTCAG
120
CCAACAAGCC TTTGCTCCCT TTGGAGAAGA ATGCTTCCAA GGTTCTTGTT GCAGGAACCC
180
ATCCTGATAA TCTGGGTTAT CAGTGTGGTG GATGGACGAT GGAATGGCAA GGATTAAGTG
240
GAAACATAAC CGTAGGAACT ACAATTCTGG AAGCTATCAA ACTAGCTGTC AGCCCCTCTA
300
CTGAAGTGGT TTATGAGCAA AATCCAGATG CTAACTATGT CAAAGGACAA GGGTTTTCAT
360
ATGCCATTGT GGTTGTGGGT GAGGCACCAT ACGCAGAAAC GTTTGGAGAC AATCTTAATT
420
TGACCATTCC CCTAGGCGGA GGGGACACGA TTAAGACGGT CTGTGGCTCC TTGAAATGCC
480
TTGTAATCTT GATATCTGGA AGGCCACTTG TTATTGAACC TTATCTTCCA TTGGTGGATC
540
GTTTT
545

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 463 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

GCAGGTCGAC ACTAGTGGAT CCAAAGAATT CGGCACGAGA AAAAACAAAT GTTAGCTAGC
60
CTAGTGATGA GCTTTACGTA TACCTGGCCT TTTATACATG GATCTGAGTT TTTATGCAGG
120
TGTAGAGCCT TTTGTTACTC TGTATCACTG GGACTTGCCA CAAGCTCTGG AGGACGAATA
180
CGGTGGATTT CGTAGCAAAA AAGTTGTGGA TGACTTTGGC ATATTCTCAG AAGAATGCTT
240
TCGTGCTTTT GGAGACCGTG TGAAGTACTG GGTAACTGTT AACGAACCGT TGATCTTCTC
300
ATATTTTTCT TACGATGTGG GGCTTCACGC ACCGGGCCGC TGTTCGCCTG GATTTGGAAA
360
CTGCACTGCG GGAAATTCAG CGACAGAGCC TTATATTGTA GCCCATAACA TGCTTCTTGC
420
ACATAGTACC GCTGTTAAAA ATATATAGCA TAAATACCCA GGG
463

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

ACACTAGTGG ATCCAAAGAA TTCGGCACGA GGCTACCATC TTCCCTCATA ATATTGGGCT
60
TGGAGCTACC AGGGATCCTG ATCTGGCTAG AAGAATAGGG GCTGCTACGG CTTTGGAAGT
120
TCGAGCTACT GGCATTCAAT ACACATTTGC TCCATGTGTT GCTGTTTGCA GAGATCCTCG
180

ATGGGGCCGC TGCTATGAGA GCTACAGTGA GGATCCAAAA ATTGTCAAGG CCATGACTGA
240
GATTATCGTT GGCCTGCAAG GGAATCCTCC TGCTAATTCT ACAAAAGGGG GGCCTTTTAT
300
AGCTGGACAG TCAAATGTTG CAGCTTGTGC TAAGCATTTT GTGGGTTATG GTGGAACAAC
360
CAAAGGTATC GATGAGAATA ATACTGTTAT CAACTATCAA GGGTTATTTC AACATTCCAA
420
ATTACCCCCA ATTTT
435

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GAATTCGGCA CGAGCCTAGA ATTCTATGGT GAAAATTGTT GGGACAAGGC TGCCCAAGTT
60
TACAAAGGAA CAGTCCCAAA TGGTTAAAGG TTCAATAGAC TATCTAGGCG TTAACCAATA
120
CACTGCTTAT TACATGTATG ATCCTAAACA ACCTAAACAA AATGTAACAG ATTACCAGAC
180
TGGACTGGAA TACAGGCTTT GCATATGCTC GCAATGGAGT GCCTATTGGA CCAAGGGCGA
240
ACTCCAATTG GCTTTACATT GTGCCTTGGG GTCTATACAA GGCCGTCACA TACGTAAAAG
300
AACACTATGG AAATCCAACT ATGATTCTCT CTGAAAATGG AATGGACGAC CTGGAAACGT
360
GACACTTCCA GCAGGACTGC ATGATACCAT CAGGGGTAAC TACTATAAAA GCTATTTGCA
420
AAATTTGATT AATGCACGTG AATGACCGGG G
451

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

CTGCTCTGCA AGCAGTACTA TGCACAGCAA GGCCTGCTTA ACTGAAAACA GAGCGCTGAG
60
CTTGAGGAAA CGCTCAAGCA TTGCTGAGGC CACCGTTTAT CTAAATAGCG CAACATAGGG
120
CTTCAGAAAA ATGGCAATGG CACAAGCATT CAGAGGCCGT GTCTTGCAAG CTGCCCGTTT
180
GCTCCGCCGC AACATTCTGC CGGAGGATAA AAGCTTTGGA TCCGCTGCTT CTCCTAGACG
240
AGCTCTTAGC CTGCTCTCAT CAAAAGCCTT CATCTCTTTC TCTGTTGAAC GGCATCGGCT
300
AGCTGCTACA AATTCAACAA TTGTGTTGCA ATCTCGAAAC TTTTCTGCAA AAGGTAAAAA
360
GACAGGACAA TCTG
374

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 457 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

GAAGAATGGA AGAGATTAAT GGTGATAACG CAGTAAGGAG GAGCTGCTTT CCTCCAGGTT
60
TCATGTTTGG GATAGCAACT TCTGCTTATC AGTGTGAAGG AGCTGCCAAC GAAGGTGGAA
120
AAGGCCCAAG CATCTGGGAC TCATTTTCAC GAACACCAGG CAAAATTCTT GATGGAAGCA
180
ACGGTGATGT AGCAGTGGAT CAGTATCATC GTTATAAGGC AGATGTAAAA CTGATGAAAG
240
ATATGGGCGT GGCTACCTAC AGATTCTCGA TTTCATGGCC TCGTATATTT CCAAAGGGAA
300
AAGGAGAGAT CAATGAGGAA GGAGTAGCCT ATTACAATAA CCTCATCAAT GAACTCCTCC
360
AGAATGGAAT CCAAGCGTCT GTCAACTTTG TTTCACTGGG ATACTCCCCA GTCTCTGGAG
420
GATGAATATG GCGGATTTCT GAGGCCAACC ATTGTGA
457

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GGTGTGATGG CAGGAATTCC AGTCCTAAGG CCATTTTGCA TCTGTTTGCT TTCAGTCTAC
60
ATGCTGCACA TTGTAGCTGC AGTAGCTTCA CCAAGGCTAG GTAGAAGCAG CTTCCCAAGG
120
GGTTTCAAAT TTGGTGCAGG GTCATCTGCT TATCAGGCGG AAGGAGCTGC TCATGAGGGT
180
GGCAAAGGCC CAAGCATTTG GGATACATTC TCCCACACTC CAGGTAAAAT CGCTGATGGG
240
AATATTGGGA TGTTGCAGTA GATCAATACC ACCGTTATAA GGAAGATGTG CAGCTTCTCA
300
AATACATGGG AATGGACGTC TATCGTTTCT CTATCTCCTG GTCACG
346

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 957 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

GAATTCGGCA CGAGAAAGCC CTAGAATTTT TTCAGCATGC TATCACAGCC CCAGCGACAA
60
CTTTAACTGC AATAACTGTG GAAGCGTACA AAAAGTTTGT CCTAGTTTCT CTCATTCAGA
120

CTGGTCAGGT TCCAGCATTT CCAAAATACA CACCTGCTGT TGTCCAAAGA AATTTGAAAT
180
CTTGCACTCA GCCCTACATT GATTTAGCAA ACAACTACAG TAGTGGGAAA ATTTCTGTAT
240
TGGAAGCTTG TGTCAACACG AACACAGAGA AGTTCAAGAA TGATAGTAAT TTGGGGTTAG
300
TCAAGCAAGT TTTGTCATCT CTTTATAAAC GGAATATTCA GAGATTGACA CAGACATATC
360
TGACCCTCTC TCTTCAAGAC ATAGCAAGTA CGGTACAGTT GGAGACTGCT AAGCAGGCTG
420
AACTCCATGT TCTGCAGATG ATTCAAGATG GTGAGATTTT TGCAACCATA AATCAGAAAG
480
ATGGGATGGT GAGCTTCAAT GAGGATCCTG AACAGTACAA AACATGTCAG ATGACTGAAT
540
ATATAGATAC TGCAATTCGG AGAATCATGG CACTATCAAA GAAGCTCACC ACAGTAGATG
600
AGCAGATTTC GTGTGATCAT TCCTACCTGA GTAAGGTGGG GAGAGAGCGT TCAAGATTTG
660
ACATAGATGA TTTTGATACT GTTCCCCAGA AGTTCACAAA TATGTAACAA ATGATGTAAA
720
TCATCTTCAA GACTCGCTTA TATTCATTAC TTTCTATGTG AATTGATAGT CTGTTAACAA
780
TAGTACTGTG GCTGAGTCCA GAAAGGATCT CTCGGTATTA TCACTTGACA TGCCATCAAA
840
AAAATCTCAA ATTTCTCGAT GTCTAGTCTT GATTTTGATT ATGAATGCGA CTTTTAGTTG
900
TGACATTTGA GCACCTCGAG TGAACTACAA AGTTGCATGT TAAAAAAAAA AAAAAAA
957

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 489 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

GCAGGTCGAC ACTAGTGGAT CCAAAGAATT CGGCACGAGA TAAGACTAAT TTTCCAGACA
60
ATCCTCCATT CCCATTCAAT TACACTGGTA CTCCACCCAA TAATACACAG GCTGTGAATG
120
GGACTAGAGT AAAAGTCCTT CCCTTTAACA CAACTGTTCA ATTGATTCTT CAAGACACCA
180
GCATCTTCAG CACAGACAGC CACCCTGTCC ATCTCCATGG TTTCAATTTC TTTGTGGTGG
240
GCCAAGGTGT TGGAAACTAC AATGAATCAA CAGATGCACC AAATTTTAAC CTCATTGACC
300
CTGTCGAGAG AAACACTGTG GGAGTTCCCA AAGGAGGTTG GGCTGCTATA AGATTTCGTG
360
CAGACAATCC AGGGGTTTGG TTCATGCACT GTCATTTGGA GGTTCACACA TCGTGGGGAC
420
TGAAAATGGC GTGGGTAGTA AAGAACGGAA AAGGGCCCAT CGATTTTCCA CCCGGGTGGG
480
TACCAGTAA
489

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 471 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

GAATTCGGCA CGAGAAAACC TTTTCAGACG AATGTTCTGA TGCTCGGCCC CGGCCAGACA
60
ACAGACATAC TTCTCACTGC CAATCAGGCT ACAGGTAGAT ACTACATGGC TGCTCGAGCA
120
TATTCCAACG GGCAAGGAGT TCCCTTCGAT AACACCACTA CCACTGCCAT TTTAGAATAC
180
GAGGGAAGCT CTAAGACTTC AACTCCAGTC ATGCCTAATC TTCCATTCTA TAACGACACC
240
AACAGTGCTA CTAGCTTCGC TAATGGTCTT AGAAGCTTGG GCTCACACGA CCACCCAGTC
300
TTCGTTCCTC AGAGTGTGGA GGAGAATCTG TTCTACACCA TCGGTTTGGG GTTGATCAAA
360
TGTCCGGGGC AGTCTTGTGG AGGTCCAACG GATCAAGATT TGCAGCAAGT ATGAATACAT
420
ATCATTTGTC CCGCAACCAC TTCTTCCAAT CCTTCAAGCT CAGCATTTTG G
471

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 338 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

GTTCGGCACT GAGAGATCCA TTTCTTTCAA TGTTGAGACA GTGAGTAGTA TTAGTTTGAT
60
ATCTCTTTCA GGAATATATC GTGCTTGCAG GATCTTTAGT TTCTGCAACA ATGTCGTTGC
120
AATCAGTGCG TCTATCTTCT GCTCTCCTTG TTTTGCTACT AGCATTTGTT GCTTACTTAG
180
TTGCTGTAAC AAACGCAGAT GTCCACAATT ATACCTTCAT TATTAGAAAG AGACAGTTAC
240
CAGGCTATGC AATAAGCGTA TAATCGCCAC CGTCAATGGC AGCTACCAGG CCCAACTATT
300
CATGTACGTG ATGGAGACGT TGTTAATTAT CAAAGCTT
338

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1229 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

AGAGAAATAA TTATATTTGT AAATTTAAGT CTACGTTTAT TAAAAAACTA CAACCCTAAA
60
TGCAGGAGAA AAAACAAGCA TGCTGTCTAC TGAAGCTTAC AAATCAAATC CCTGCGATAT
120
GTCTTTTCTC GTGCCGAATT CGGCACGAGA AGATCTTGGT TCGAGTCTCT CAGCTCTCTC
180
CAAAGGAATT TTGTGGGTCA TTTGCAGGTG AAGACACCAT GGTGAAGGCT TATCCCACCG
240
TAAGCGAGGA GTACAAGGCT GCCATTGACA AATGCAAGAG GAAGCTCCGA GCTCTCATTG
300

CAGAGAAGAA CTGTGCGCCG ATCATGGTTC GAATCGCATG GCACAGCGCT GGGACTTACG
360
ATGTCAAGAC CAAGACCGGA GGGCCCTTCG GGACGATGAG ATATGGGGCC GAGCTTGCCC
420
ACGGTGCTAA CAGTGGTCTG GACATCGCAG TTAGGCTCCT GGAGCCAATC AAGGAACAGT
480
TCCCCATAAT CACCTATGCT GACCTTTATC AGTTGGCTGG TGTGGTGGCT GTTGAAGTGA
540
CCGGGGGACC TGACATTCCG TTCCATCCTG GAAGAGAAGA CAAGCCTGAG CCTCCAGAAG
600
AAGGCCGCCT TCCTGATGCT ACAAAAGGAC CTGATCATCT GAGGGATGTT TTTGGTCACA
660
TGGGGTTGAA TGATAAGGAA ATTGTGGCCT TGTCTGGTGC CCACACCTTG GGGAGATGCC
720
ACAAGGAGAG ATCTGGTTTT GAAGGACCAT GGACCTCTAA CCCCCTTATC TTTGACAACT
780
CTTACTTCAC AGAGCTTGTG ACTGGAGAGA AGGAAGGCCT GCTTCAGTTG CCATCTGATA
840
AGGCACTGCT TGCTGATCCT AGTTTTGCAG TTTATGTTCA GAAGTATGCA CAGGACGAAG
900
ACGCTTTCTT TGCTGACTAT GCGGAAGCTC ACCTGAAGCT TTCTGAACTT GGGTTTGCTG
960
ATGCGTAGAT TCATACCTTC TGCAGAGACA ATTCCTTGCT AGATAGCTTC GTTTTGTATT
1020
TCATCTAATC TTTTCGATTA TATAGTCACA TAGAAGTTGG TGTTATGCGC CATAGTGATA
1080
CTTGAACCTA CATGTTTTTG AAAAGTATCG ATGTTCTTTA AAATGAACAT TGAATACAAC
1140
ATTTTGGAAT CTGGTTGTGT TCTATCAAGC GCATATTTTA ATCGAATGCT TCGTTCCTGT
1200
TAAAAAAAAA AATAAAATAA AAAAAAAAA
1229

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1410 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

GAAGATGGGG CTGTGGGTGG TGCTGGCTTT GGCGCTCAGT GCGCACTATT GCAGTCTCAG
60
GCTTACAATG TGGTAAGTTC AAGCAATGCT ACTGGGAGTT ACAGTGAGAA TGGATTGGTG
120
ATGAATTACT ATGGGGACTC TTGCCCTCAG GCTGAAGAGA TCATTGCTGA ACAAGTACGC
180
CTGTTGTACA AAAGACACAA GAACACTGCA TTCTCATGGC TTAGAAATAT TTTCCATGAC
240
TGTGCTGTGG AGTCATGTGA TGCATCGCTT CTGTTGGACT CAACAAGGAA CAGCATATCA
300
GAAAAGGACA CTGACAGGAG CTTCGGCCTC CGCAACTTTA GGTATTTGGA TACCATCAAG
360
GAAGCCGTGG AGAGGGAGTG CCCCGGGGTC GTTTCCTGTG CAGATATACT CGTTCTCTCT
420
GCCAGAGATG GCGTTGTATC GTTGGGAGGA CCATACATTC CCCTGAAGAC GGGAAGAAGA
480
GATGGACGGA AGAGCAGAGC AGATGTGGTG GAGAATTACC TGCCCGATCA CAATGAGAGC
540
ATCTCCACTG TTCTGTCTCG CTTCAAAGCC ATGGGAATCG ACACCCGTGG GGTTGTTGCA
600
CTGCTGGGGG CTCACAGCGT GGGGAGGACT CACTGCGTGA AGCTGGTGCA CAGGCTGTAC
660

CCGGAAGTAG ATCCGACACT GGACCCTGGG CACGTGGAGC ACATGAAGCA CAAGTGCCCG
720
GACGCGATCC CCAACCCGAA GGCAGTGCAG TATGTGCGGA ACGACCGGGG AACGCCTATG
780
AAGCTGGACA ACAACTACTA CGTGAACCTG ATGAACAACA AGGGGCTCCT AATAGTGGAC
840
CAGCAACTGT ATGCAGATTC GAGGACCAGG CCGTATGTGA AGAAGATGGC AAAAAGCCAG
900
GAATACTTCT TCAAATACTT CTCCCGGGCG CTCACCATCC TCTCTGAGAA CAATCCTCTC
960
ACCGGCGCTC GAGGAGAAAT CCGTCGGCAG TGCTCGCTCA AAAACAAATT CCACACAAAA
1020
AGCAAGCGTT GAGCGATAGC TCAATGCCGC AGTGGTGGGA GTGATAGCGT GATGCCACAG
1080
TGGTGGGCAT TTCATATATA AATTGCAGTT TGCGTTTTTA TTAGATAATC ATAATGGTGT
1140
GGTGTGACTA TGCCCTGCGA ATCACATCGA TGAACCACAA CCGAACCGTG GAACAGTAGG
1200
CTTATTCCCT TATGTAAGCA GAACCTTTTA TTATAAGCAA AAAAGACAAT CCTGTCTGTT
1260
ATTCTAGTAT AATTTTGTCA TCAGTTAAAG TTGCTCATCT GATAATAACT GGAAACGGTA
1320
AAATATGACA ACTACGTATC TTCTTTGGTC ATCTGATAAT AACCGGAAAC GATAAAATAT
1380
GACAACTACA TATATTCTTT AAAAAAAAAA
1410

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 687 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GTAGTTTCGT TTTACAACAA TCTCAGGTTT TGAATCTCAG AATAGTTGCG AAAGGAAGCG
60
ATGACGAAGT ACGTGATCGT TAGCTCCATT GTGTGTTTCT TTGTATTTGT TTCTGCGTGC
120
ATAATTTCTG TCAATGGATT AGTTGTCCAT GAAGATGATC TGTCAAAGCC TGTGCATGGG
180
CTTTCGTGGA CATTTTATAA GGACAGTTGC CCCGACTTGG AGGCCATAGT GAAATCGGTA
240
CTTGAGCCGG CGTTGGACGA AGATATCACT CAGGCCGCAG GCTTGCTGAG ACTTCATTTC
300
CATGACTGTT TTGTGCAGGG TTGCGATGGG TCCGTGTTGC TGACAGGAAC TAAAAGAAAC
360
CCCAGTGAGC AACAGGCTCA GCCAAACTTA ACACTAAGAG CCCGGGCCTT GCAGCTGATC
420
GACGAAATTA AAACCGCTGT AGAAGCTAGC TGCAGTGGGG TTGTAACTTG TGCAGACATT
480
CTGGCTTTGG CTGCTCGTGA CTCCGTCCGC TCAGGAGGCC CAAAATTTCC AGTACCACTT
540
GGCCGCAGAG ATAGCCTAAA GTTTGCCAGT CAATCCGTAG TTCTCGCCAA TATACCAACT
600
CCAACTTTAA ATTTGACACA GCTGATGAAC ATTTTTGGCT CCAAAGGATT CAGTTTGGCC
660
GAAATGGTTG CTCTTCAGGT GGCACAC
687

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 688 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GTAGTTTCGT TTTACAACAA TCTACAGGTT TTGAATCTCA GAATAGTTGC GAAAGGAAGC
60
GATGACGAAG TACGTGATCG TTAGCTCCAT TGTATGTTTC TTTGTATTTG TTTCTGCGTG
120
CATAATTTCT GTCAATGGAT TAGTTGTCCA TGAAGATGAT CTGTCAAAGC CTGTGCATGG
180
GCTTTCGTGG ACATTTTATA AGGACAGTTG CCCCGACTTG GAGGCCATAG TGAAATCGGT
240
ACTTGAGCCG GCGTTGGACG AAGATATCAC TCAGGCCGCA GGTTGCTGAG ACTTCATTTC
300
CATGACTGTT TTGTGCAGGG TTGCGATGGG TCCGTGTTGC TGACAGGAAC TAAAAGAAAC
360
CCCCGAGTGA GCAACAGGCT CAGCCAAACT TAACACTAAG AGCCCGGGCC TTGCAGCTGA
420
TCGACGAAAT TAAAACCGCT GTAGAAGCTA GCTGCAGTGG GGTTGTAACT TGTGCAGACA
480
TTCTGGCTTT GGCTGCTCGT GACTCCGTCG CTCAGGAGGC CCAAAATTTC CAGTACCACT
540
TGGCCGCAGA GATAGCCTAA AGTTTGCCAG TCAATCCGTA GTTCTCGCCA ATATACCAAC
600
TCCAACTTTA AATTTGACAC AGCTGATGAA CATTTTTGGC TCCAAAGGAT TCAGTTTGGC
660
CGAAATGGTT GCTCTTCAGG TGGCACAC
688

Claims:

1. An isolated DNA sequence comprising a nucleotide sequence selected from the group consisting of

(a) sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88;

(b) complements of the sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88;

(c) reverse complements of the sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88;

(d) reverse sequences of the sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and

(e) sequences having at least about a 99% probability of being the same as a sequence of (a) - (d) as measured by computer algorithm FASTA.

2. A DNA construct comprising a DNA sequence according to claim 1.

3. A transgenic cell comprising a DNA construct according to claim 2.

4. A DNA construct comprising, in the 5'-3' direction:

(a) a gene promoter sequence,

(b) an open reading frame coding for at least a functional portion of an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA; and

(c) a gene termination sequence.

5. The DNA construct of claim 4 wherein the open reading frame is in a sense orientation.

6. The DNA construct of claim 4 wherein the open reading frame is in an antisense orientation.

7. The DNA construct of claim 4, wherein the gene promoter sequence and gene termination sequences are functional in a plant host.

8. The DNA construct of claim 4, wherein the gene promoter sequence provides for transcription in xylem.

9. The DNA construct of claim 4 further comprising a marker for identification of transformed cells.

10. A DNA construct comprising, in the 5`-3` direction:
(a) a gene promoter sequence,
(b) a non-coding region of a gene coding for an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA; and
(c) a gene termination sequence.

11. The DNA construct of claim 10 wherein the non-coding region is in a sense orientation.

12. The DNA construct of claim 10 wherein the non-coding region is in an antisense orientation.

13. The DNA construct of claim 10, wherein the gene promoter sequence and gene termination sequences are functional in a plant host.

14. The DNA construct of claim 10, wherein the gene promoter sequence provides for transcription in xylem.

15. A transgenic plant cell comprising a DNA construct, the DNA construct comprising, in the 5'-3' direction:

(a) a gene promoter sequence;

(b) an open reading frame coding for at least a functional portion of an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA; and

(c) a gene termination sequence.

16. The transgenic plant cell of claim 15 wherein the open reading frame is in a sense orientation.

17. The transgenic plant cell of claim 15 wherein the open reading frame is in an antisense orientation.

18. The transgenic plant cell of claim 15 wherein the DNA construct further comprises a marker for identification of transformed cells.

19. A plant comprising a transgenic plant cell according to claim 15, or fruit or seeds thereof.

20. The plant of claim 19 wherein the plant is a woody plant.

21. The plant of claim 20 wherein the plant is selected from the group consisting of eucalyptus and pine species.

22. A transgenic plant cell comprising a DNA construct, the DNA construct comprising, in the 5'-3' direction:

(a) a gene promoter sequence;

(b) a non-coding region of a gene coding for an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA; and

(c) a gene termination sequence.

23. The transgenic plant cell of claim 22 wherein the non-coding region is in a sense orientation.

24. The transgenic plant cell of claim 22 wherein the non-coding region is in an antisense orientation.

25. A plant comprising a transgenic plant cell according to claim 22, or fruit or seeds thereof.

26. The plant of claim 25 wherein the plant is a woody plant.

27. The plant of claim 26, wherein the plant is selected from the group consisting of eucalyptus and pine species.

28. A method for modulating the lignin content of a plant comprising stably incorporating into the genome of the plant a DNA construct comprising, in the 5'-3' direction:

(a) a gene promoter sequence;

(b) an open reading frame coding for at least a functional portion of an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA; and

(c) a gene termination sequence.

29. The method of claim 28 wherein the plant is selected from the group consisting of eucalyptus and pine species.

30. The method of claim 28 wherein the open reading frame is in a sense orientation.

31. The method of claim 28 wherein the open reading frame is in an antisense orientation.

32. A method for modulating the lignin content of a plant comprising stably incorporating into the genome of the plant a DNA construct comprising, in the 5'-3' direction:

(a) a gene promoter sequence:

(b) a non-coding region of a gene coding for an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA: and

(c) a gene termination sequence.

33. The method of claim 32 wherein the non-coding region is in a sense orientation.

34. The method of claim 32 wherein the non-coding region is in an antisense orientation.

35. The method of claim 32 wherein the plant is a woody plant.

36. The method of claim 35, wherein the plant is selected from the group consisting of eucalyptus and pine species.

37. A method for producing a plant having altered lignin structure comprising:

(a) transforming a plant cell with a DNA construct comprising, in the 5`-3` direction, a gene promoter sequence, an open reading frame coding for at least a functional portion of an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA, and a gene termination sequence to provide a transgenic cell;

(b) cultivating the transgenic cell under conditions conducive to regeneration and mature plant growth.

38. The method of claim 37 wherein the open reading frame is in a sense orientation.

39. The method of claim 37 wherein the open reading frame is in an antisense orientation.

40. The method of claim 37 wherein the plant is a woody plant.

41. The method of claim 40 wherein the plant is selected from the group consisting of eucalyptus and pine species.

42. A method for producing a plant having altered lignin structure comprising:

(a) transforming a plant cell with a DNA construct comprising, in the 5`-3` direction, a gene promoter sequence, a non-coding region of a gene coding for an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA, and a gene termination sequence to provide a transgenic cell:

(b) cultivating the transgenic cell under conditions conducive to regeneration and mature plant growth.

43. The method of claim 42 wherein the non-coding region is in a sense orientation.

44. The method of claim 42 wherein the non-coding region is in an antisense orientation.

45. The method of claim 42 wherein the plant is a woody plant.

46. The method of claim 45 wherein the plant is selected from the group consisting of eucalyptus and pine species.

47. A method of modifying the activity of an enzyme in a plant comprising stably incorporating into the genome of the plant a DNA construct including

(a) a gene promoter sequence;

(b) an open reading frame coding for at least a functional portion of an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA; and

(c) a gene termination sequence.

48. The method of claim 47 wherein the open reading frame is in a sense orientation.

49. The method of claim 47 wherein the open reading frame is in an antisense orientation.

50. A method of modifying the activity of an enzyme in a plant comprising stably incorporating into the genome of the plant a DNA construct including

(a) a gene promoter sequence;

(b) a non-coding region of a gene coding for an enzyme encoded by a nucleotide sequence selected from the group consisting of sequences recited in SEQ ID NO: 3, 13, 16-70, 72-88 and sequences having at least about a 99% probability of being the same as a sequence of SEQ ID NO: 3, 13, 16-70, 72-88 as measured by computer algorithm FASTA; and

(c) a gene termination sequence.

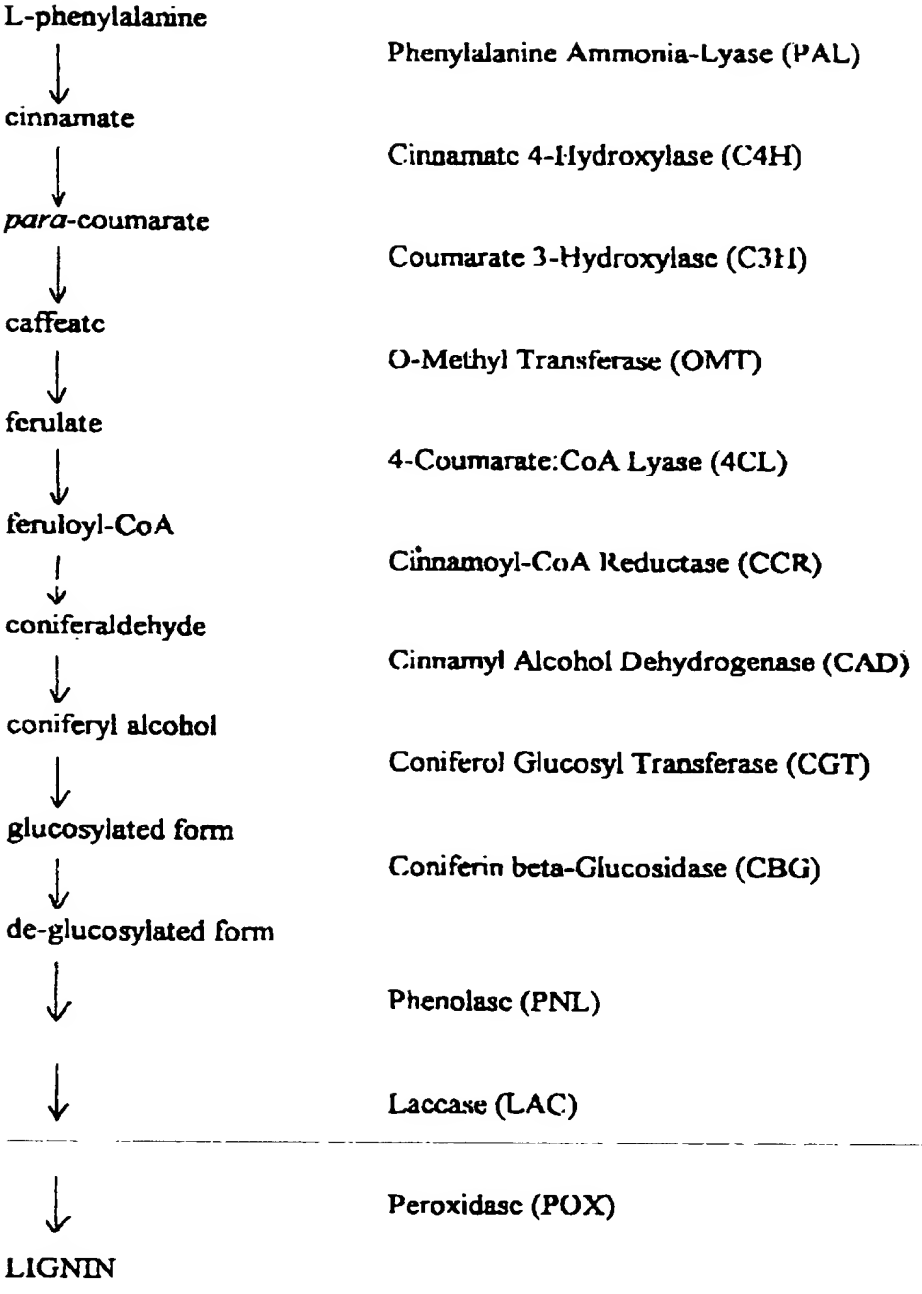
51. The method of claim 50 wherein the non-coding region is in a sense orientation.

52. The method of claim 50 wherein the non-coding region is in an antisense orientation.

53. The method of claim 50 wherein the plant is a woody plant.

54. The method of claim 53 wherein the plant is selected from the group consisting of eucalyptus and pine species.

FIG. 1



PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/82, 15/11, 15/63, A01H 1/00, 5/00 | **A3**

(11) International Publication Number: **WO 99/24561**

(43) International Publication Date: 20 May 1999 (20.05.99)

(21) International Application Number: PCT/US98/24138

(22) International Filing Date: 12 November 1998 (12.11.98)

(30) Priority Data:
08/969,046 12 November 1997 (12.11.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/969,046 (CIP)
Filed on 12 November 1997 (12.11.97)

(71) Applicant *(for all designated States except US)*: BOARD OF CONTROL OF MICHIGAN TECHNOLOGICAL UNIVERSITY [US/US]; 1400 Townsend Drive, Houghton, MI 49931 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: CHIANG, Vincent, Lee, C. [US/US]; 1104 Birch Street, Hancock, MI 49930 (US). TSAI, Chung-Jui [-/US]; 1510 Anthony Street #34, Hancock, MI 49930 (US). HU, Wen-Jing [-/US]; 8299 Cambridge Drive #1003, Houston, TX 77054 (US).

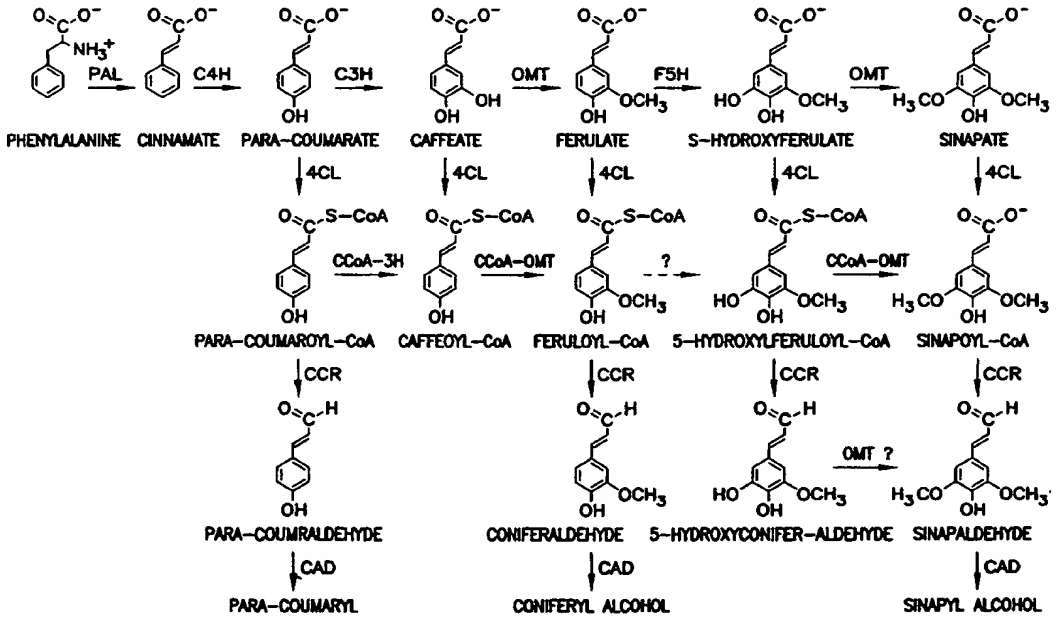
(74) Agent: VIKSNINS, Ann, S.; Schwegman, Lundberg, Woessner & Kluth, P.O. Box 2938, Minneapolis, MN 55402 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report:
15 July 1999 (15.07.99)

(54) Title: GENETIC ENGINEERING OF LIGNIN BIOSYNTHESIS IN PLANTS



(57) Abstract

The invention pertains to altering a lignin pathway 4-coumarate Co-enzyme A ligase (4CL) in plants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

Interna[illegible] Application No
PCT/[illegible] 98/24138

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N15/11 C12N15/63 A01H1/00 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 A01H C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DOUGLAS, C.J. ET AL.: "Exonic sequences are required for elicitor and light activation of a plant defense gene, but promoter sequences are sufficient for tissue specific expression" THE EMBO JOURNAL, vol. 10, no. 7, July 1991, pages 1767-1775, XP002100277 see page 1770, sec. column; page 1771,first column,lines 7-13 and sec. column, lines 11-14;page 1773, sec. column, lines 6-8 and last paragraph contind. on page 1774, lines 1-5; page 1774, first column, lines 24-26, sec. column, first paragraph;	1,3,4,6, 7,10, 12-15, 26, 28-31, 33,42-44
A	page 1771, first column, lines 11-12	47-50

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 April 1999	05.05.99

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Alt, G

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
X	KAJITA, S. ET AL.: "Alterations in the biosynthesis of lignin in transgenic plants with chimeric gens for 4-coumarate:coenzyme A ligase" PLANT CELL PHYSIOLOGY, vol. 37, no. 7, 1996, pages 957-965, XP002065207 page 958, second column, lines 1-12; Figure 2;page 960, sec. column, sec. paragraph;page 961, sec. column; 962, sec.column; page 963, first column, first paragraph	1,2, 4-16, 18-26, 36-42
A	ZHANG, X.-H. AND CHIANG, V.L.: "Molecular cloning of 4-coumarate:coenzym A Ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood" PLANT PHYSIOLOGY, vol. 113, January 1997, page 65-74 XP002100278 see page 70, first column, lines 33-36	28, 36-41,43
A	UHLMANN,A. AND EBEL, J.: "Molecular cloning and expression of 4-coumarate:coenzyme A ligase, an enzyme involved in the resistance response of soybean (Glycine max L.) against pathogen attack" PLANT PHYSIOLOGY, XP002100279 see the abstract	1
A	LEE, D. ET AL.: "The Arabidopsis thaliana 4-coumarate:CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA" PLANT MOLECULAR BIOLOGY, vol. 28, 1995, pages 871-884, XP002100280 see page 876, second column, lines 16-20	47-50
P,X	WO 98 11205 A (GENESIS RESEARCH & DEVELOPMENT CORPORATION LIMITED) 19 March 1998 see Example 4	3-7, 10-14, 18-20, 22,24, 25,36,37

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
T	HU, W.J. ET AL.: "Compartmentalized expression of two structuraly and functionally distinct 4-coumarate:CoA ligase genes in aspen (Populus tremoides)" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 95, April 1998, pages 5407-5412, XP002100281 see the whole document	1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 98/24138

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees

INTERNATIONAL SEARCH REPORT

International Application No. PCT/ US 98/24138

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-50 (1-17)

   Method for altering the growth characteristics of a plant by incorporating into the genome a DNA molecule comprising a nucleotide sequence encoding 4-coumarate-Co-enzyme A ligase and corresponding plants.

2. Claims: 18-25

   Method for altering the characterisitic of a plant, the characteristic selected from the group of accelerated growth, reduced lignin content, altered lignin structure, increased disease resistance and increased cellulose content, by genetically down-regulating the enzyme 4-coumarate Co-enzyme A ligase and corresponding plants

3. Claims: 26-28, 42, 43, 45

   A DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and directing expression to the xylem.

4. Claims: 26, 27, 29, 42, 44

   DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and directing expression to epidermal tissue

5. Claims: 30-35

   Method of imparting disease resistance to a plant tissue by introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase and corresponding plants and seeds

6. Claims: 36, 37

   Method for altering the lignin content in a plant by introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase and corresponding plants

7. Claims: 38, 39

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Method for altering the cellulose content in a plant by introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase and corresponding plants

8. Claims: 40, 41

Method for altering the lignin structure in a plant by introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase and corresponding plants

9. Claims: 47-50

Method for enhancing the root growth of a plant by incorporating into the genome of the plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase and corresponding plants

The ISA considers that the present claims do not relate to one invention or a group of inventions so linked as to form a single general inventive concept as required by Rule 13.1 PCT. The reasoning is as follows:
Currently, the inventive concept linking all claims can be considered as methods for altering the growth characteristics of a plant by incorporating into the genome of a plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase or regulatory parts thereof.

This concept is however known from Kajita, S. et al., Plant Cell Physiology, vol. 37, no. 7 (1996), pages 957-965. The document discloses that the introduction of 4-coumarate:coenzyme A ligase (4CL) chimeric sense and antisense genes into tobacco caused the reduction of 4CL acitivty. The observed effects were that the cell walls of the xylem tissue in stems were brown, that the molecular structure of lignin in the colored cell walls was different from that of control plants and that the lignin content was reduced.

Thus, since the above defined inventive concept is not novel, the application is considered as being directed to nine different inventions which are not linked by corresponding special technical features. The specific features are:
1. Claims 1-17: Incorporation into the genome a nucleotide sequence encoding 4-coumarate-Co-enzyme A ligase for altering the growth characteristics.

2. Claims 18-25: Genetically down regulating the enzyme 4-coumarate Co-enzyme A ligase for altering the characteristic of a plant, the characteristic selected from the group of accelarated growth, reduced lignin content, altered lignin structure, increased disease

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

resistance and increased cellulose content.

3. Claims 26-28, 42, 43, 45: A DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and directing expression to the xylem.

4. Claims 26, 27, 29, 42, 44: DNA molecule comprising a DNA segment comprising a transcriptional regulatory region of a plant 4-coumarate Co-enzyme A ligase and expression cassette containing said segment and directing expression to epidermal tissue.

5. Clais 30-35, 46: Introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase for imparting disease resistance.

6. Claims 36, 37: Introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase for altering the lignin content.

7. Claims 38, 39: Introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase for latering the cellulose content.

8. Claims 40, 41: Introducing an expression cassette comprising a recombinant DNA molecule comprising a nucleotide sequence encoding a 4-coumarate Co-enzyme A ligase for altering the lignin structure.

9. Claims 47-50: Incorporating into the genome of the plant a recombinant DNA molecule comprising a nucleotide sequence encoding 4-coumarate Co-enzyme A ligase for enhancing root growth.

INTERNATIONAL SEARCH REPORT

mation on patent family members

International Application No
PCT/US 98/24138

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9811205 A	19-03-1998	US 5850020 A	15-12-1998
		AU 4403697 A	02-04-1998